

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

### **IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

	Type	Hits	Search Text	DBs	Time Stamp	Comments	Error Definition
49	IS&R	2	("6381608").PN.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/18 17:32		
50	BRS	336	"edit decision"	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/18 17:32		
51	BRS	82	"edit decision" same (first or last)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/18 17:34		
52	BRS	24	("edit decision" same (first or last)) and integrat\$3	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/18 17:34		
53	BRS	922	cut with dissolve	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 15:08		
54	BRS	31	(cut with dissolve) and EDL	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 15:08		
55	BRS	1	first with last with decision with edit	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 17:34		
56	BRS	244719	first same last saem decision same edit	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 17:35		
57	BRS	466	(first same last saem decision same edit) and edl	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 17:35		
58	BRS	5	((first same last saem decision same edit) and edl) and recalculat\$3	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:22		
59	IS&R	2438	(707/102).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:23		
60	IS&R	1488	(707/101).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:23		
61	IS&R	3367	(707/104.1).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:23		
62	IS&R	161	(710/30).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:23		
63	IS&R	979	(711/100).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:23		
64	IS&R	910	(386/52).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:23		
65	IS&R	176	(386/55).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:24		
66	IS&R	456	(345/723).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:24		
67	IS&R	12	(345/724).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:24		
68	IS&R	40	(345/726).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:24		
69	IS&R	72	(345/730).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:24		
70	IS&R	40	(345/732).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:24		
71	IS&R	358	(348/722).CCLS.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/08/19 20:24		

	Errors
49	0
50	0
51	0
52	0
53	0
54	0
55	0
56	0
57	0
58	0
59	0
60	0
61	0
62	0
63	0
64	0
65	0
66	0
67	0
68	0
69	0
70	0
71	0

IEEE HOME | SEARCH IEEE | SHOP | WEB ACCOUNT | CONTACT IEEE



Membership Publications/Services Standards Conferences Careers/Jobs

Welcome  
United States Patent and Trademark Office
[Help](#) [FAQ](#) [Terms](#) [IEEE Peer Review](#)
[Quick Links](#)

» Search Results

## Welcome to IEEE Xplore®

- ☐ Home
- ☐ What Can I Access?
- ☐ Log-out

## Tables of Contents

- ☐ Journals & Magazines
- ☐ Conference Proceedings
- ☐ Standards

## Search

- ☐ By Author
- ☐ Basic
- ☐ Advanced

## Member Services

- ☐ Join IEEE
- ☐ Establish IEEE Web Account
- ☐ Access the IEEE Member Digital Library

## IEEE Enterprise

- ☐ Access the IEEE Enterprise File Cabinet

Your search matched **1** of **1062489** documents.A maximum of **500** results are displayed, **15** to a page, sorted by **Relevance** in **Descending** order.

## Refine This Search:

You may refine your search by editing the current search expression or entering a new one in the text box.

☐ Check to search within this result set

## Results Key:

**JNL** = Journal or Magazine   **CNF** = Conference   **STD** = Standard1 **Elefant and Kanguru: hardware control panels dedicated to disk servers for slow motion, editing and multiplay applications***Grimaldi, J.L.; Bougard, M.;*

Broadcasting Convention, 1997. International , 12-16 Sept. 1997

Pages:268 - 271

[\[Abstract\]](#)   [\[PDF Full-Text \(268 KB\)\]](#)   [IEE CNF](#)

## Print Format

[Home](#) | [Log-out](#) | [Journals](#) | [Conference Proceedings](#) | [Standards](#) | [Search by Author](#) | [Basic Search](#) | [Advanced Search](#) | [Join IEEE](#) | [Web Account](#) | [New this week](#) | [OPAC Linking Information](#) | [Your Feedback](#) | [Technical Support](#) | [Email Alerting](#) | [No Robots Please](#) | [Release Notes](#) | [IEEE Online Publications](#) | [Help](#) | [FAQ](#) | [Terms](#) | [Back to Top](#)

Copyright © 2004 IEEE — All rights reserved



US Patent &amp; Trademark Office

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: ☒ The ACM Digital Library ☐ The Guide

+edit +frame +integrat\*



decision list +

THE ACM DIGITAL LIBRARY


[Feedback](#) [Report a problem](#) [Satisfaction survey](#)
Terms used [edit](#) [frame](#) [integrat](#)

Found 13 of 160 searched out of 160.

Sort results by

relevance

[Save results to a Binder](#)

Display results

expanded form

[Search Tips](#)☐ Open results in a new window[Try an Advanced Search](#)[Try this search in The ACM Guide](#)

Results 1 - 13 of 13

Relevance scale ☐ ☐ ☐ ☐ ☐1 Integrated video archive tools

Rune Hjelsvold, Stein Langørgen, Roger Midstraum, Olav Sandst

January 1995 **Proceedings of the third ACM international conference on Multimedia**Full text available: [htm\(45.77 KB\)](#)Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)**Keywords:** applications, browsing, content-based retrieval, digital libraries, video databases2 Future directions in desktop video

T. Heidmnn, M. MacKay, G. MacNicol, F. Wray

July 1989 **ACM SIGGRAPH Computer Graphics , ACM SIGGRAPH 89 Panel Proceedings**, Volume 23 Issue 5Full text available: [pdf\(2.66 MB\)](#)Additional Information: [full citation](#), [abstract](#), [index terms](#)

Good morning. My name is Tim Heidmann and I'd like to welcome you all to this panel, which is entitled Future Directions in Desktop Video, and I'd especially like to thank all you people who stayed up a little late on Thursday night to come to this panel. It's really good to see you all out there.

I've gotten word that this panel is being transcribed. They're putting together a booklet, so they're taking the slides and the stills from the videos and all the things that we're ...

3 CVEPS - a compressed video editing and parsing system

Jianhao Meng, Shih-Fu Chang

February 1997 **Proceedings of the fourth ACM international conference on Multimedia**Full text available: [pdf\(1.38 MB\)](#)Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)4 DVI—a digital multimedia technology

G. David Ripley

July 1989 **Communications of the ACM**, Volume 32 Issue 7Full text available: [pdf\(4.55 MB\)](#)Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

A digital presentation technology that manages anything from text to full-motion video has the potential of expanding the usefulness of personal computers, while rendering them less intimidating.

5 A semi-automatic approach to home video editing

Andreas Girgensohn, John Boreczky, Patrick Chiu, John Doherty, Jonathan Foote, Gene Golovchinsky, Shingo Uchihashi, Lynn Wilcox  
November 2000 **Proceedings of the 13th annual ACM symposium on User interface software and technology**

Full text available:  pdf(1.06 MB) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

**Keywords:** automatic video clip extraction, video analysis, video editing, video exploration

6 Authoring Support: Designing annotation before it's needed

Frank Nack, Wolfgang Putz  
October 2001 **Proceedings of the ninth ACM international conference on Multimedia**

Full text available:  pdf(1.16 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

This paper considers the automated and semi-automated annotation of audiovisual media in a new type of production framework, A4SM (Authoring System for Syntactic, Semantic and Semiotic Modelling). We present the architecture of the framework and outline the underlying XML-Schema based content description structures of A4SM. We then describe tools for a news and demonstrate how video material can be annotated in real time and how this information can not only be used for retrieval but also can be ...

**Keywords:** MPEG-7, XML Schema, automated annotation, news production, semantic networks

7 Zodiac: a history-based interactive video authoring system

Tzi-cker Chiueh, Tulika Mitra, Anindya Neogi, Chuan-Kai Yang  
September 1998 **Proceedings of the sixth ACM international conference on Multimedia**

Full text available:  pdf(1.10 MB) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

8 Direct manipulation of temporal structures in a multimedia application framework

P. Ackermann  
October 1994 **Proceedings of the second ACM international conference on Multimedia**

Full text available:  pdf(883.13 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

The design of complex multimedia documents presents new challenges to authoring systems, because spatial and temporal features should be visualized and made accessible in an intuitive and direct-manipulative way. In this study, multimedia presentations are considered as hierarchical compositions of time objects that define serial or parallel synchronization of the inserted media objects. Media composition hierarchies support automatic temporal layout mechanisms. They are integrated into an ...

9 Passive capture and structuring of lectures

Sugata Mukhopadhyay, Brian Smith  
October 1999 **Proceedings of the seventh ACM international conference on Multimedia (Part 1)**

Full text available:  pdf(2.15 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Despite recent advances in authoring systems and tools, creating multimedia presentations remains a labor-intensive process. This paper describes a system for automatically constructing structured multimedia documents from live presentations. The automatically


produced documents contain synchronized and edited audio, video, images, and text. Two essential problems, synchronization of captured data and automatic editing, are identified and solved.

**Keywords:** audio/video capture, educational technology, matching

#### 10 Device reservation in audio/video editing systems

David P. Anderson

May 1997 **ACM Transactions on Computer Systems (TOCS)**, Volume 15 Issue 2

Full text available:  pdf(297.15 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#), [review](#)


What fraction of disks and other shared devices must be reserved to play an audio/video document without dropouts? In general, this question cannot be answered precisely. For documents with complex and irregular structure, such as those arising in audio/video editing, it is difficult even to give a good estimate. We describe three approaches to this problem. The first, based on long-term average properties of segments, is fast but imprecise: it underreserves in some cases and overreserves i ...

**Keywords:** admission control, edit decision list, quality of service, reservation

#### 11 Session 12: interfacing stored media II: Creating music videos using automatic media analysis

Jonathan Foote, Matthew Cooper, Andreas Girgensohn

December 2002 **Proceedings of the tenth ACM international conference on Multimedia**

Full text available:  pdf(1.19 MB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)

We present methods for automatic and semi-automatic creation of music videos, given an arbitrary audio soundtrack and source video. Significant audio changes are automatically detected; similarly, the source video is automatically segmented and analyzed for suitability based on camera motion and exposure. Video with excessive camera motion or poor contrast is penalized with a high unsuitability score, and is more likely to be discarded in the final edit. High quality video clips are then automat ...



**Keywords:** audio analysis, music video, video analysis, video editing

#### 12 Special issue on word sense disambiguation: Introduction to the special issue on word sense disambiguation: the state of the art

Nancy Ide, Jean Véronis

March 1998 **Computational Linguistics**, Volume 24 Issue 1

Full text available:

 pdf(3.44 MB)   
[Publisher Site](#)

Additional Information: [full citation](#), [references](#), [citations](#)

#### 13 Xanalogical structure, needed now more than ever: parallel documents, deep links to content, deep versioning, and deep re-use

Theodor Holm Nelson

December 1999 **ACM Computing Surveys (CSUR)**

Full text available:  pdf(787.72 KB)

Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2004 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:  [Adobe Acrobat](#)  [QuickTime](#)  [Windows Media Player](#)  [Real Player](#)





US006571051B2

(12) **United States Patent**  
**Savoie**

(10) **Patent No.:** **US 6,571,051 B2**  
(45) **Date of Patent:** **May 27, 2003**

(54) **EDITING IMAGE DATA**

(75) **Inventor:** Charles Savoie, Quebec (CA)

(73) **Assignee:** Autodesk Canada Inc., Quebec (CA)

(\*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) **Appl. No.:** 09/126,687

(22) **Filed:** Jul. 30, 1998

(65) **Prior Publication Data**

US 2002/0076198 A1 Jun. 20, 2002

(30) **Foreign Application Priority Data**

Aug. 1, 1997 (GB) ..... 97 16 248

(51) **Int. Cl.<sup>7</sup>** ..... H04N 5/93

(52) **U.S. Cl.** ..... 386/55

(58) **Field of Search** ..... 386/46, 52, 55,  
386/65; 345/328, 723

(56) **References Cited**

**U.S. PATENT DOCUMENTS**

4,121,253 A	10/1978	McCoy	
4,855,834 A	8/1989	Cawley et al.	
4,954,898 A	9/1990	Nakata	
5,184,222 A	2/1993	Yanagisawa	
5,191,645 A	3/1993	Carlucci et al.	
5,237,648 A	8/1993	Mills et al.	
5,339,393 A	8/1994	Duffy et al.	
5,353,391 A	10/1994	Cohen et al.	
5,359,712 A	10/1994	Cohen et al.	
5,388,197 A	2/1995	Rayner	
5,404,316 A	4/1995	Klingler et al.	
5,442,744 A	8/1995	Piech et al.	
5,488,433 A *	1/1996	Washino et al.	386/55
5,515,110 A	5/1996	Alig et al.	
5,568,275 A	10/1996	Norton et al.	
5,758,180 A	5/1998	Duffy et al.	

5,760,767 A	6/1998	Shore et al.	
5,781,188 A	7/1998	Amiot et al.	
5,786,814 A	7/1998	Moran et al.	
5,926,603 A *	7/1999	Tanaka et al.	386/52
5,930,445 A *	7/1999	Peters et al.	386/52
5,982,364 A	11/1999	Beckwith	
5,999,173 A	12/1999	Ubillos	
6,016,380 A *	1/2000	Norton	386/55
6,034,679 A *	3/2000	McGraph	345/328

**FOREIGN PATENT DOCUMENTS**

EP	0438061 A1	7/1991	
EP	0625781 A3	11/1994	G11B/27/028
EP	0625781 A2	11/1994	G11B/27/028
EP	2306750 A	5/1997	
GB	2219170 A	11/1989	
GB	2272127 A	5/1994	
GB	2280778 A	2/1995	
GB	2300535 A	11/1996	
GB	2312078 A	10/1997	G11B/27/034
GB	2295482 B	12/1997	G11B/27/031

\* cited by examiner

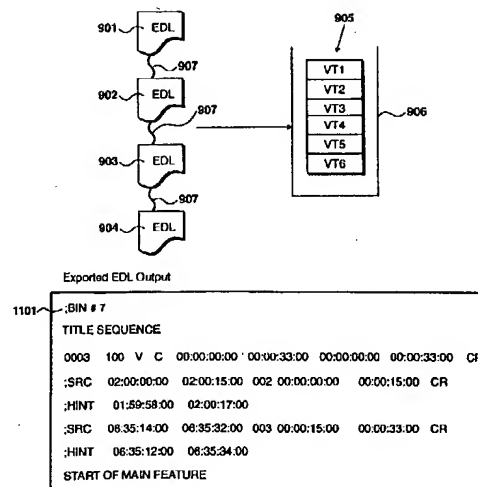
*Primary Examiner*—Huy Nguyen

(74) *Attorney, Agent, or Firm*—Gates & Cooper LLP

(57) **ABSTRACT**

An image data editing system is arranged to capture source image clips at full definition in response to edit decision lists. The system conforms the captured material to produce output image clips in response to the edit decisions defined in the list used to capture the source material. Association identifiers are applied to many edit decision lists and the capturing process is configured to capture all source material for all of the edit decision lists having the same association identifier. Each associated edit decision list may be conformed independently and manipulations may be performed with reference to the captured source material. However, when many output reels are required derived from a collection of source material, it is only necessary to capture the source material once.

**50 Claims, 11 Drawing Sheets**





US 20020076198A1

(19) **United States**(12) **Patent Application Publication**  
**SAVOIE**(10) **Pub. No.: US 2002/0076198 A1**(43) **Pub. Date: Jun. 20, 2002**(54) **EDITING IMAGE DATA****Publication Classification**(76) **Inventor: CHARLES SAVOIE, MONTREAL (CA)**(51) **Int. Cl.<sup>7</sup> ..... H04N 5/76; G11B 27/00**(52) **U.S. Cl. .... 386/55; 386/64**

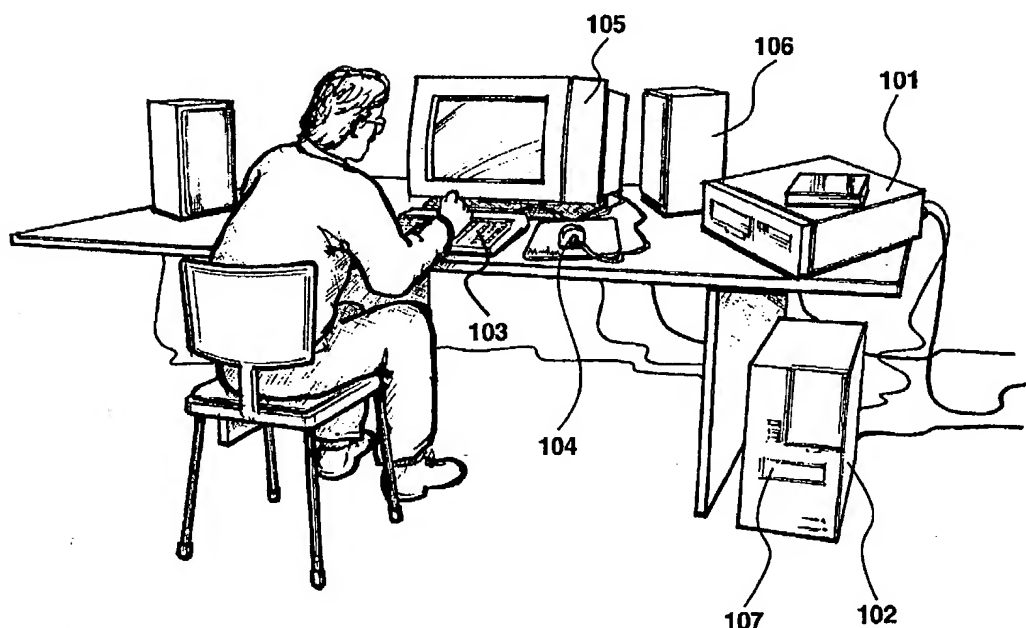
Correspondence Address:

**GATES & COOPER LLP****HOWARD HUGHES CENTER****6701 CENTER DRIVE WEST, SUITE 1050****LOS ANGELES, CA 90045 (US)**(57) **ABSTRACT**(\*) **Notice:** This is a publication of a continued prosecution application (CPA) filed under 37 CFR 1.53(d).

An image data editing system is arranged to capture source image clips at full definition in response to edit decision lists. The system conforms the captured material to produce output image clips in response to the edit decisions defined in the list used to capture the source material. Association identifiers are applied to many edit decision lists and the capturing process is configured to capture all source material for all of the edit decision lists having the same association identifier. Each associated edit decision list may be conformed independently and manipulations may be performed with reference to the captured source material. However, when many output reels are required derived from a collection of source material, it is only necessary to capture the source material once.

(21) **Appl. No.: 09/126,687**(22) **Filed: Jul. 30, 1998**(30) **Foreign Application Priority Data**

Aug. 1, 1997 (GB) ..... GB 97 16 248.1

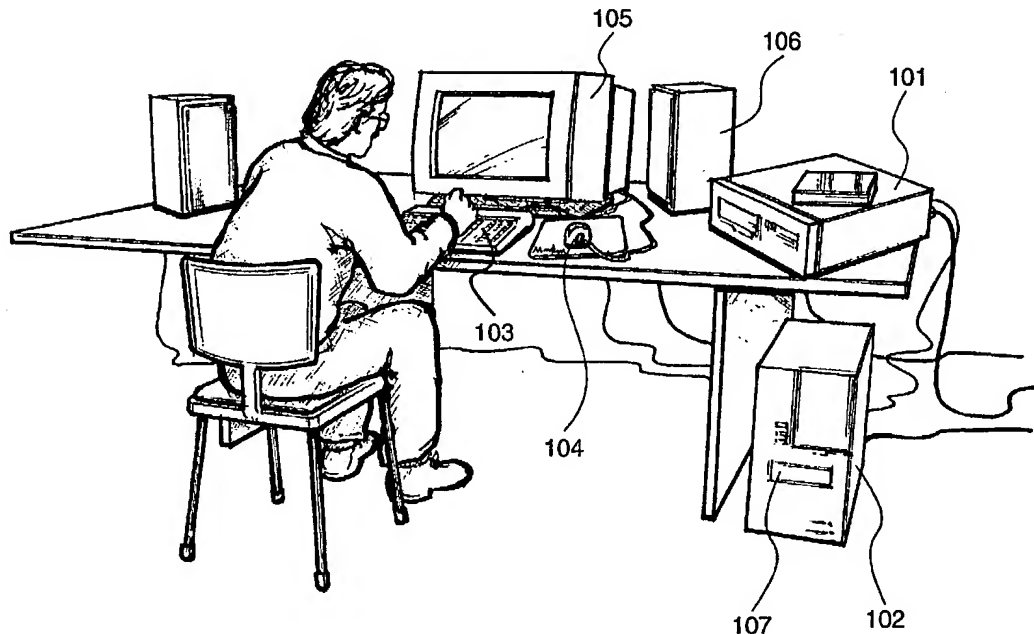




US 20020044759A1

(19) **United States**(12) **Patent Application Publication** (10) **Pub. No.: US 2002/0044759 A1**  
COLLIN et al. (43) **Pub. Date: Apr. 18, 2002**(54) **ON-LINE EDITING AND DATA CONVEYING  
MEDIA FOR EDIT DECISIONS**(22) **Filed: Aug. 13, 1997****Publication Classification**(76) **Inventors: DANIEL COLLIN, QUEBEC (CA);  
CHARLES SAVOIE, QUEBEC (CA)**(51) **Int. Cl.<sup>7</sup> ..... H04N 5/76; G11B 27/00**(52) **U.S. Cl. .... 386/55; 386/64****Correspondence Address:****GATES & COOPER LLP****HOWARD HUGHES CENTER****6701 CENTER DRIVE WEST, SUITE 1050****LOS ANGELES, CA 90045 (US)**(57) **ABSTRACT**

Frames of image data (FIG. 5) are conformed from source material in response to an edit decision list (FIG. 2). Some of the image data is modified in an on-line environment (FIG. 3) and an output tape is produced including said modified frames. An output edit decision list (FIG. 7) is generated with conventional editing references (SRC, HINT) to the output tape. In addition, reference to the original source material are also included, encoded as comments.

( \* ) **Notice:** This is a publication of a continued prosecution application (CPA) filed under 37 CFR 1.53(d).(21) **Appl. No.: 08/910,818**



US006625385B2

(12) **United States Patent**  
Collin et al.

(10) **Patent No.:** US 6,625,385 B2  
(45) **Date of Patent:** Sep. 23, 2003

(54) **ON-LINE EDITING AND DATA CONVEYING  
MEDIA FOR EDIT DECISIONS**

(75) **Inventors:** Daniel Collin, Quebec (CA); Charles  
Savole, Quebec (CA)

(73) **Assignee:** Autodesk Canada Inc., Quebec (CA)

(\*) **Notice:** Subject to any disclaimer, the term of this  
patent is extended or adjusted under 35  
U.S.C. 154(b) by 0 days.

(21) **Appl. No.:** 08/910,818

(22) **Filed:** Aug. 13, 1997

(65) **Prior Publication Data**

US 2002/0044759 A1 Apr. 18, 2002

(51) **Int. Cl.<sup>7</sup>** ..... H04N 5/93

(52) **U.S. Cl.** ..... 386/52; 386/55

(58) **Field of Search** ..... 386/46, 52, 53,  
386/54, 55; 345/328; 703/23

(56) **References Cited**

#### U.S. PATENT DOCUMENTS

5,404,316 A \* 4/1995 Klingler et al. .... 345/328  
5,469,270 A \* 11/1995 Yamamoto ..... 386/55

5,488,433 A \* 1/1996 Washino et al. .... 386/55  
5,568,275 A \* 10/1996 Norton et al. .... 386/55  
5,584,006 A \* 12/1996 Reber et al. .... 386/55  
5,649,046 A \* 7/1997 Stewart et al. .... 386/52  
5,649,171 A \* 7/1997 Craven et al. .... 703/23  
5,754,730 A \* 5/1998 Windrem et al. .... 386/124  
5,771,330 A \* 6/1998 Takano et al. .... 386/52  
6,016,380 A \* 1/2000 Norton ..... 386/52

#### FOREIGN PATENT DOCUMENTS

EP 0438299 A2 \* 7/1991

\* cited by examiner

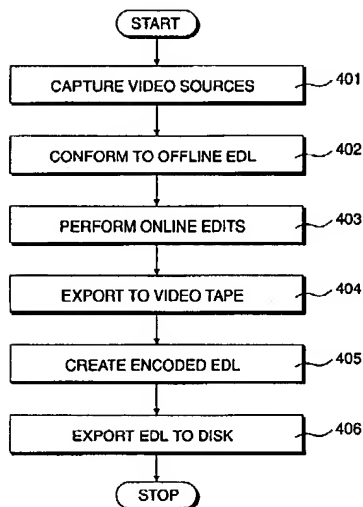
*Primary Examiner*—Huy Nguyen

(74) *Attorney, Agent, or Firm*—Gates & Cooper LLP

(57) **ABSTRACT**

Frames of image data (FIG. 5) are conformed from source material in response to an edit decision list (FIG. 2). Some of the image data is modified in an on-line environment (FIG. 3) and an output tape is produced including said modified frames. An output edit decision list (FIG. 7) is generated with conventional editing references (SRC, HINT) to the output tape. In addition, reference to the original source material are also included, encoded as comments.

26 Claims, 7 Drawing Sheets



#### Encoded (Online) EDL Output

```

TITLE SEQUENCE
0003 100 V C 00000000 00003300 00000000 00000000 CR
701 :SRC 02000000 02001500 002 00000000 00001500 CR
702 :HINT 01595800 02001700
703 :SRC 08351400 08353200 003 00001500 00003300 CR
704 :HINT 08351200 08353400
START OF MAIN FEATURE
  
```

# CVEPS - A Compressed Video Editing and Parsing System

Jianhao Meng and Shih-Fu Chang

Department of Electrical Engineering &

Center for Image Technology for New Media

Columbia University

New York, NY 10027, USA

{jmeng,sfchang}@itnm.columbia.edu

<http://www.itnm.columbia.edu/~jmeng,-sfchang>

## ABSTRACT

Processing digital video directly in the compressed domain has many advantages in terms of storage efficiency, speed, and video quality. We have developed a compressed video editing and parsing system (CVEPS) with advanced video indexing and manipulation functions. The video parsing tools support automatic extraction of key visual features, e.g., scene cuts, transitional effects, camera operations (zoom/pan), shape and trajectories of prominent moving objects. These visual features are used for efficient video indexing, retrieval and browsing. The editing tools allow users to perform useful video compositing functions and special visual effects typically seen in video production studios. We contrast our compressed-domain approach with traditional decode-process-reencode approach with quantitative and/or qualitative performance comparison. We also present a client-server network based CVEPS implementation.

## KEYWORDS

Compressed domain video manipulation, client-server network based video editing, video content analysis, video indexing.

## 1. INTRODUCTION

Digital video is an essential component of new media applications. It demands special technical support in processing, communication, and storage. This paper investigates innovative compressed-domain technologies for compressed video manipulation, indexing, and browsing, in order to support various multimedia applications such as real-time video production and video digital library.

We present a *Compressed Video Editing and Parsing System*, CVEPS, using a unique compressed-domain approach

which offers many great benefits [6,7]. First, implementation of the same manipulation algorithms in the compressed domain will be much cheaper than that in the uncompressed domain because the data rate is highly reduced in the compressed domain (e.g., a typical 20:1 to 50:1 compression ratio for MPEG). Second, given most existing images and videos stored in the compressed form, the specific manipulation algorithms can be applied to the compressed streams without full decoding of the compressed images/videos. Lastly, because that full decoding and re-encoding of video are not necessary, we can avoid the extra quality degradation that usually occurs in the reencoding process. We have shown that for MPEG compressed video editing, the speed performance can be improved by more than 60 times and the video quality can be improved by about 3-4 dB if we use the compressed-domain approach rather than the traditional decode-edit-reencode approach [15].

In order to allow users to manipulate compressed video directly, two types of functionalities are required (1) key content browsing and search, (2) compressed video editing. The former allows users to efficiently browse through or search for key content of the video without decoding and viewing the entire video stream. The key content refers to the key frames in video sequences, prominent video objects and their associated visual features (motion, shape, color, and trajectory), or special reconstructed video models for representing video content in a video scene. The second type of functionalities, video editing, allow users to manipulate the object of interest in the video stream without full decoding. One example is to cut and paste any arbitrary segments from existing video streams and produce a new video stream which conforms to the valid compression format. Other examples include special visual effects typically used in video production studios.

This paper describes system components and specific proposed compressed-domain algorithms for achieving the above functionalities in CVEPS. The primary compression standard used is MPEG (MPEG1 and MPEG2). Most of our techniques are applicable to generally encoded MPEG streams with different parameter settings such as constant or variable bitrate, different frequency of I, P, B frames etc. Our scene change detection techniques assume the use of interframe coded frames (i.e. P or B). However, the underlying

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee.

ACM Multimedia 96, Boston MA USA

© 1996 ACM 0-89791-871-1/96/11 ...\$3.50

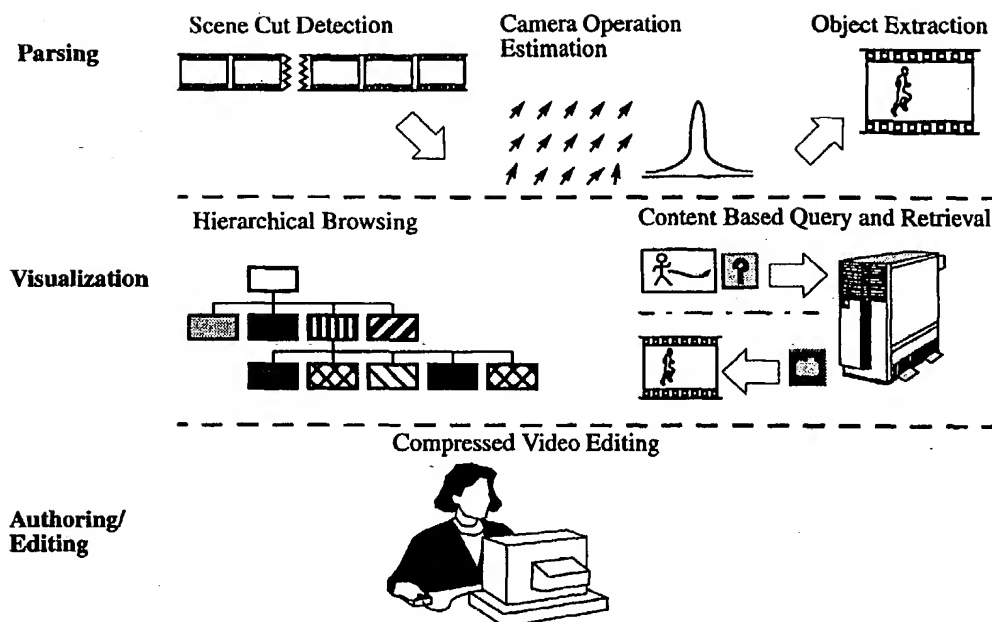


FIGURE 1. CVEPS System Overview

ing approach and techniques are general enough to be applied to other video compression standards (e.g. those using transform coding and/or interframe motion compensation). This paper is organized as the following. Section 2 discusses related work. Section 3 provides system overview for CVEPS. Section 4 presents our compressed-domain techniques for parsing MPEG video to extract visual features. Section 5 describes algorithms for compressed video editing. Section 6 discusses system design issues, followed by conclusion at the end.

## 2. RELATED WORK

Video indexing and manipulation has emerged as an active research area. Much work has been reported by several research groups, some of which also explored the compressed-domain approach. But there are no existing systems that provide integrated solutions for both video manipulation and video indexing. To this end, our prior work has presented techniques for manipulation of both compressed image and video [7,8], compressed image feature extraction [6], and video scene analysis using MPEG streams [15].

For scene cut detection in the spatial domain, Smoliar and Zhang proposed color histogram comparison [22] and Shahraray used a block-based match and motion estimation algorithm [19]. In the compressed domain (Motion JPEG video), comparison of DCT coefficients of selected blocks from each JPEG frame was used to detect the scene cuts [4]. We detect scene cuts in motion compensated video sequences such as MPEG. Distribution of motion vectors is used for detecting direct scene cuts and the variance of DCT DC coefficients is used for detecting transitional scene cuts [14]. After the scene cuts are found, video shots can be browsed with the clustering algorithms proposed in [24].

Within each shot, camera operation and moving objects are important visual features. In spatial domain, finding parameters of an affine matrix and constructing a mosaic image from a sequence of video images was addressed by Sawhney *et al* [18]; searching for object appearance and using them in video indexing was proposed by Nagasaka *et al* [16]. In compressed domain, detecting camera operations (zoom, pan) using motion vectors had been discussed in [2,25]. Both [2,25] used a simple 3 parameter model with the assumption that the camera panning is very small and focal length is very long. The two restrictions make the algorithms not suitable for general video processing. Object motion tracking in MPEG video was also discussed by Dimitrova *et al* [9], however, camera operations were not taken into consideration for object motion recovery. We use a 6-parameter affine transform model and the least squares (LS) method to estimate camera operation parameters. With the estimated camera parameters we further recover the local object motion from the global motion.

Video indexing using finite state models for parsing and retrieval of specific domain video, such as news video, was discussed by Smoliar *et al* [22]. Hampapur *et al* [11] proposed feature based video indexing scheme, which uses low level machine derivable indices to map into the set of application specific video indices. Our goal is to extract a rich set of visual features associated with the scenes and individual objects from the compressed video to enable content based query, and allow for integration with domain knowledge for derivation of higher-level semantics.

To manipulate image and video sequences, a resolution independent video language (Rivl) was proposed by Swartz and Smith [23]. Although Rivl utilized group of pictures (GOPs) level direct copying whenever possible for "cut and

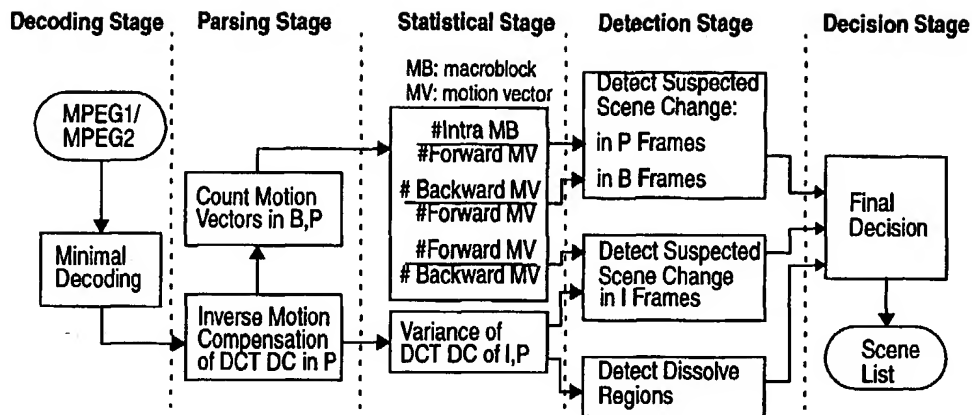


FIGURE 2. The Compressed Domain Scene Cut Detection Algorithm

paste" operations on MPEG video, Rivl did not use compressed domain approach at the frame level and macroblock level for special effects editing (see Section 5.2). Most video effects in Rivl were done by decoding each frame to pixel domain and applying image library routines. Also the rate control problems due to editing of constant bitrate video was not addressed by Rivl.

### 3. SYSTEM OVERVIEW

The CVEPS system consists of three major modules: Parsing, Visualization and Authoring, see Figure 1. In the Parsing module, MPEG compressed video is first broken into shot segments. Within each shot, camera operation parameters are estimated. Then moving objects are detected and their shape and trajectory features are extracted. In the Visualization module, the scene cut output list and the camera zoom/pan information are used to extract key frames for representing each video shot. The key frames can be browsed with the hierarchical video scene browser [26]. Our content-based image query system, VisualSEEk [20] and WebSEEk [21], are used to index and retrieve key frames or video objects based on their visual features and spatial layout. In the Authoring module, we provide tools for cutting/pasting of arbitrary MPEG video segments and adding special effects such as dissolve, key, masking and motion effects (described in more details later).

### 4. PARSING OF MPEG VIDEO

#### 4.1 Scene Cut Detection In Compressed Domain

Within a video shot, consecutive frames have high temporal correlation. In MPEG video, this correlation can be characterized by the ratio of the number of backward motion vectors (or intracoded macroblocks) versus the number of forward motion vectors in B (or P) frames. For example, when a direct scene cut occurs on a P-frame, most macroblocks will be intracoded (i.e., no interframe prediction). We calculate the motion vector ratios for every B/P frame and use local adaptive thresholds to detect the peak values.

To detect the transitional scene cut such as dissolve, we use the fact that the variance of the pixel intensity of each frame in the dissolve region shows an approximated parabolic curve [3]. For MPEG video, we use the DCT DC values to approximate the pixel intensity. We are able to successfully detect long dissolves in sequences without high motion. Short dissolves with high motion are trickier and often treated as direct scene cuts.

Figure 2 shows the block diagram of our scene cut detection algorithm. MPEG video is minimally decoded and parsed to get the motion vector counts and DCT DC coefficients. This involves simple parsing of the MPEG streams and does not need any intensive computation. In the Statistical Stage, three ratios are calculated for detecting direct scene cuts in P, B, and I frames, respectively; variance of DCT DC coefficients are calculated from I and P frames for detecting dissolve curves. The peaks of ratios and the dissolve curve are found in the Detection Stage. Finally, duplicated cuts are eliminated before returning a list of scenes.

We have tested our algorithms on several bitstreams from classic movies and CNN news. Table 1 shows the results of a 10 minutes CNN news (unconstrained content) with 19931 frames, Group of Pictures (GOP) size 15, one I or P frame for every two B frames, and frame size 352 pixels by 240 pixels. For the direct scene cuts, we detected 54 out of 59 correctly; the 7 false alarms were mainly caused by a shot including the strobe motion special effect (refer to Section 5.2.4); the 5 missed cuts were due to similar dark background of the two shots. For transitional effects, we detected 19 out of 21 correctly; the false alarms and misses in the transitional scene cut detection were mainly due to our light-weight implementation which skipped B frames.

TABLE 1. Scene Cut Detection Results

	Direct Scene Cuts	Transitional Scene Cuts
Manual	59	21
Detected	54	19
Missed	5	2
False Alarm	7	8

#### 4.2 Camera Operation Parameters Estimation

Within a shot, low level visual features such as camera zoom/pan and moving objects are useful information for video indexing. We estimate the camera zoom and pan with a 6-parameter affine transform model [5] using the motion vectors from the MPEG compressed stream.

The motion vectors in MPEG are usually generated by block matching: finding a block in the reference frame so that the mean square error is minimized. Although the motion vectors do not represent the true optical flow, it is still good in most cases to estimate the camera parameters in sequences that do not contain large dark or uniform regions.

When the distance between the object/background and the camera is large, it is usually sufficient to use a 6 parameter affine transform to describe the global motion of the current frame,

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix} \cdot \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \end{bmatrix}^T \quad (1)$$

where  $(x, y)$  is the coordinate of a macroblock in the current frame,  $\begin{bmatrix} u & v \end{bmatrix}^T$  is the motion vector associated with that macroblock,  $\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \end{bmatrix}^T$  is the affine transform vector. We denote  $U$  for  $\begin{bmatrix} u & v \end{bmatrix}^T$ ,  $X$  for  $\begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix}$ , and  $\hat{a}$  for  $\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \end{bmatrix}^T$ .

Given the motion vector for each macroblock, we find the global parameter using the Least Squares (LS) estimation, that is to find a set of parameter  $\hat{a}$  to minimize the error between the motion vectors estimated in (1) and the actual motion vectors obtained from the MPEG stream [25].

$$S(\hat{a}) = \sum_x \sum_y [(\hat{u}_{xy} - u_{xy})^2 + (\hat{v}_{xy} - v_{xy})^2] \quad (2)$$

where  $\begin{bmatrix} \hat{u} & \hat{v} \end{bmatrix}^T$  is the estimated motion vector. To solve for  $\hat{a}$ , set the first derivative of  $S(\hat{a})$  to 0, then we get

$$\begin{bmatrix} N & A & B \\ A & C & E \\ B & E & D \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} \text{ and } \begin{bmatrix} N & A & B \\ A & C & E \\ B & E & D \end{bmatrix} \cdot \begin{bmatrix} a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} \quad (3)$$

where,

$$\begin{aligned} N &= \sum_x \sum_y 1, \quad A = \sum_x \sum_y x, \quad B = \sum_x \sum_y y, \\ C &= \sum_x \sum_y x^2, \quad D = \sum_x \sum_y y^2, \quad E = \sum_x \sum_y x \cdot y, \\ U_1 &= \sum_x \sum_y u_{xy}, \quad U_2 = \sum_x \sum_y u_{xy} \cdot x, \quad U_3 = \sum_x \sum_y u_{xy} \cdot y, \\ V_1 &= \sum_x \sum_y v_{xy}, \quad V_2 = \sum_x \sum_y v_{xy} \cdot x, \quad V_3 = \sum_x \sum_y v_{xy} \cdot y. \end{aligned}$$

All summations are computed over all valid macroblocks whose motion vectors survive after the nonlinear noise reduction process. After the first LS estimation, motion vectors that have large distance from the estimated ones are filtered out before a second LS estimation. The estimation process is iterated several times to refine the accuracy.

#### 4.3 Moving Object Detection and Tracking

After the global camera parameters  $\hat{a}$  is found, we may recover the object motion by applying the global motion compensation. If an object located at  $(x, y)$  in the current frame has a local motion  $M = \begin{bmatrix} m_x & m_y \end{bmatrix}^T$  from  $(x_0, y_0)$  to  $(x_1, y_1)$  in the reference frame with motion vector  $U$ , then  $U + M = X \cdot \hat{a}$ , see Figure 3. That means the local object motion can be recovered from motion vectors provided that  $\hat{a}$  is known,

$$M = X \cdot \hat{a} - U \quad (4)$$

This is the global motion compensation (GMC). For motion vectors of the background, GMC will give mostly 0. For motion vectors of the foreground moving objects, GMC will reveal the local motion of objects, see Figure 4(b).

Moving objects are detected by thresholding the magnitude of the local motion followed by simple morphological operations to delete small false objects and to fill noisy spots.

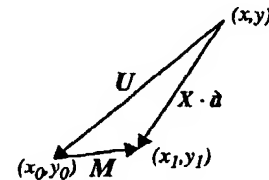


FIGURE 3. Relation among global motion  $X \cdot \hat{a}$ , local motion  $M$  and net displacement  $U$



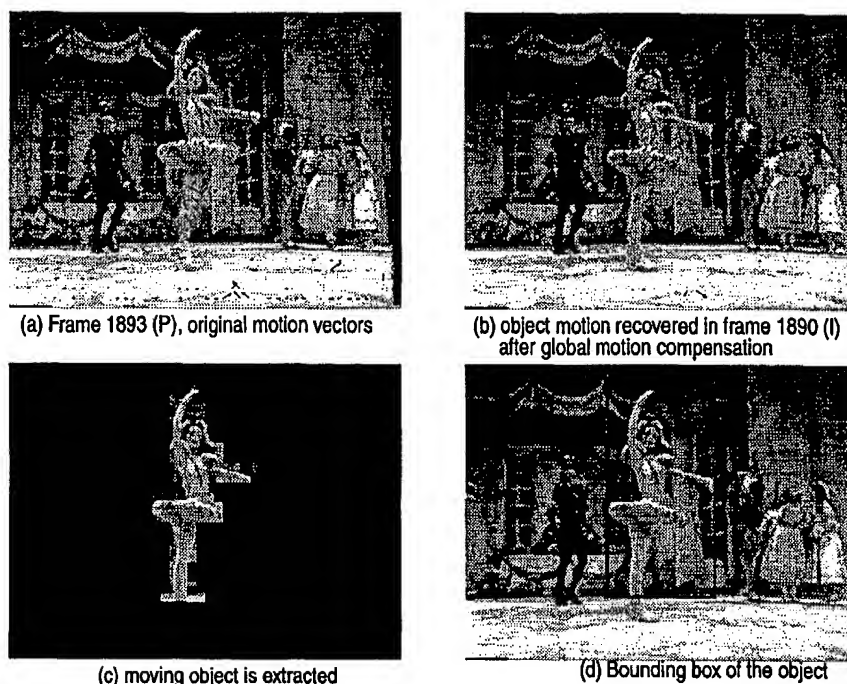


FIGURE 4. Camera Parameter and Moving Object Detection

See Figure 4(c) for extracted moving object. The DCT coefficients of the moving object are extracted for query purpose. The outermost points of the object are used to form a bounding box. The location and size of the bounding boxes are saved for later browsing and indexing, see Figure 4(d).

To track the moving objects throughout a video shot, we first select a reference frame where the moving object is initially detected. Secondly, we obtain the centroid of each moving object by taking the first moment of the object's shape. Thirdly, we map the centroid of each object onto the reference frame using the global camera parameters  $\hat{a}$ . When tracking multiple objects, color and texture of the object can be used to distinguishing them. The motion trajectory of each moving object is formed by repeatedly mapping the centroid until the object has stopped or moved out of the picture or the next scene comes. Finally, filters such as a median filter are used to smooth out the trajectories.

Visual features of the extracted objects, such as color, textures, and shape, can be used to provide content-based visual query of these and associated video scenes.

## 5. COMPRESSED VIDEO EDITING

Based on the source material, we classify video editing into two stages: the production stage and the post-production stage. The production stage editing are based on original analog or digital footages from cameras. At this level, sophisticated hardware is usually used to guarantee the ease of editing and the highest possible video quality. Commercially available digital video systems such as AVID, Media100 and D-Vision etc., currently use the Motion JPEG

compression [17]. The compression ratio varies from 3:1 to about 10:1. With the latest technology, high bandwidth bus technology will make uncompressed video editing possible. The output video from the production stage will be eventually converted to more heavily compressed bitstreams (e.g. MPEG2) for broadcasting or storage.

At the post-production stage, the users will retrieve the MPEG bitstreams according to their needs and perform desired editing. Post-production video editing shall not be available only to users that have sophisticated video hardware. We develop the CVEPS using a pure software and compressed domain approach particularly for this purpose.

We will discuss technical issues of editing MPEG video such as frame type conversion, maintaining bitrate integrity and algorithms for creating common special effects in the compressed domain.

### 5.1 Basic Editing Functions: Cut and Paste MPEG Video

When cutting and pasting several MPEG video segments to create a new sequence, a straightforward way is to decode all the segments and re-encode. This method is computation intensive, and the output picture will suffer generation loss multiple times.

We apply the basic editing functions directly in the compressed domain. Figure 5 illustrates a scenario of cutting two arbitrary segments from the middle of two separate video streams and merging them to form a new compressed video stream.

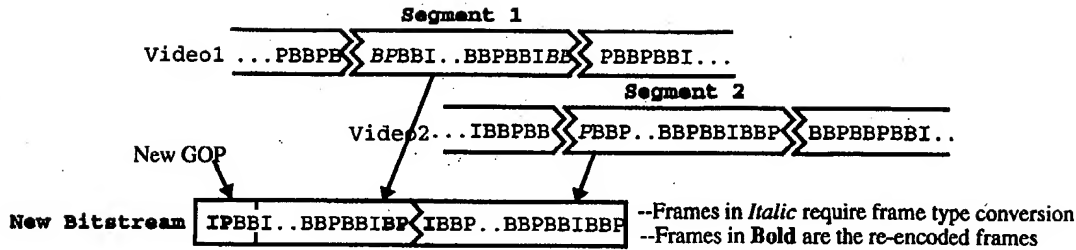


FIGURE 5. Cut and Paste MPEG bitstreams in the compressed domain.

### 5.1.1 Issue I — Frame Type Conversion

The MPEG video consists of GOP units. Each GOP starts with an I frame. We only need to re-encode few frames which are out of the GOP boundary at the beginning or ending part of the segments. The newly created GOP may have a different size, but it is still conformable to the MPEG format. Details of the frame type conversion may be found in [15]. After type conversion, each segment is independently decodable and can be pasted together back to back to form a new sequence. Figure 5 shows cutting out segment 1 and 2 at arbitrary location to form a new bitstream. The beginning few frames of a segment is re-encoded to form a shorter new GOP.

### 5.1.2 Issue II — Decoder Video Buffer Control

For constant bitrate MPEG video, the MPEG encoder solves the rate control problem with the “virtual buffer” [12,13], a simulation module of the decoder buffer. Before quantizing each macroblock, it sets the reference value of the quantization parameter based on the fullness of the “virtual buffer.”

When cutting and pasting arbitrary segments from different compressed video streams of the same bitrate, the integrity of the original rate control mechanism is lost. For example, Figure 6 (a) shows the video buffer occupancy after connecting four segments. The video buffer size is 1Mbits. Each segment consists of 49 frames, starts with an I frame and

ends with an I frame. The video buffer decreases to a very low level after the first I frame of Seg3. When Seg4 is pasted, the buffer starts to have the underflow problem.

The overflow problem can be easily solved by stuffing zero bits at the end of a slice or a picture whenever the buffer reaches a very high level. The underflow problem can be solved by inserting a synthetic transitional GOP [15] which has a lower average bitrate than normal GOPs or by applying rate shaping algorithm [10] to reduce the bitrate of the boundary I/P frames.

### 5.2 Extended Editing Functions: Special Effects in the Compressed Domain

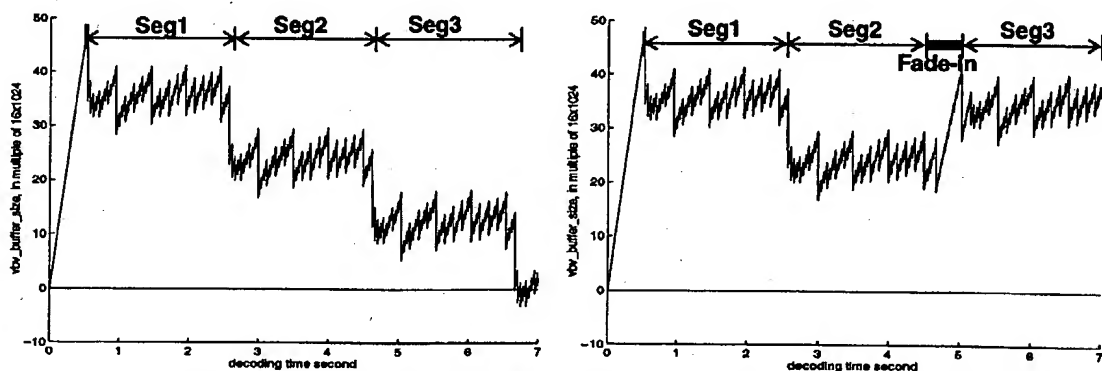
In addition to the basic editing function “cut and paste”, several special visual effects can be created in the compressed domain. For I frames, the basic compression component is the Discrete Cosine Transform (DCT), which we denote as

$$F(u, v) = DCT(f(x, y)) \quad (5)$$

Basic linear operations like the intensity addition and scaling can done as follows [7],

$$DCT(f_1(x, y) + f_2(x, y)) = F_1(u, v) + F_2(u, v) \quad (6)$$

$$DCT(\alpha \cdot f(x, y)) = \alpha \cdot F(u, v) \quad (7)$$



(a) Decoder video buffer underflows when pasting segments together. (b) With the proposed synthetic fade-in connecting Seg2 and Seg3, buffer remains normal.

FIGURE 6. Connecting MPEG video segments in the compressed domain.

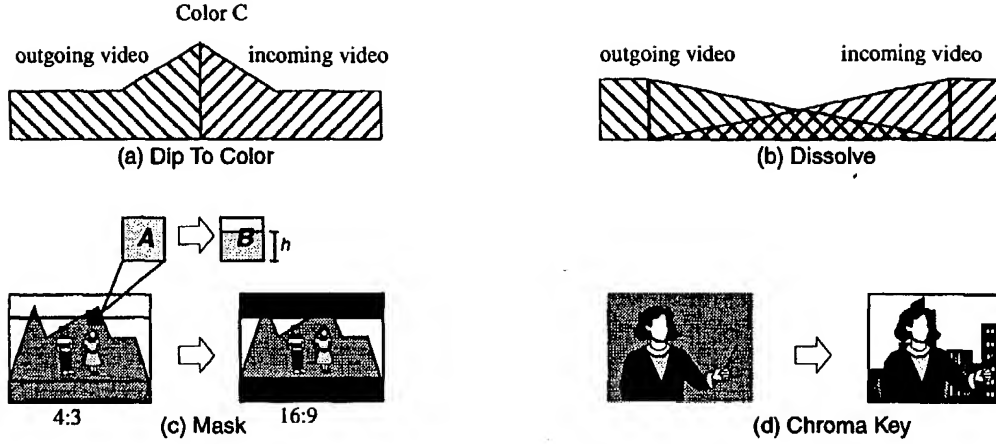


FIGURE 7. Some Special Visual Effects

Algorithms for other operations such as spatial scaling, translation, and filtering in DCT domain can be found in [7]. Usually, the DCT of the output video,  $Y$ , can be obtained by linear matrix operations of the input DCT,  $P_i$ , as follows

$$Y = \sum_i W_i \cdot P_i \cdot H_i \quad (8)$$

where  $H_i$  and  $W_i$  are special filter coefficient matrices in the DCT domain. For motion compensated B and P frames, the compressed-domain manipulation functions can be implemented in two ways. First, in [7,8], we have proposed transform-domain techniques to convert B and P frames to intraframe DCT coefficients, on which the above techniques can be readily applied. An alternative is to keep the B/P structure (i.e., DCT of residual errors and motion vectors) and develop algorithms directly utilizing these data. The following are some examples of typically used editing functions such as *Blend*, *Film*, *Key*, *Motion*, and *Wipe* etc. [1], some of which are illustrated in Figure 7.

### 5.2.1 Blend Effects

Blend effects are generally two-channel effects: to create a transitional connection between two video segments. Two commonly used ones are: dip to color and dissolve.

#### Dip to color

Fades from the outgoing video to black, white, or any color and then fades to the incoming video. Since the outgoing and incoming video do not overlap, this effect is achieved by modifying the DCT coefficients in outgoing and incoming video frames. The normalized color level increment  $\Delta I_k$  is added to the DCT DC of each macroblock,

$$\Delta I_k = \frac{N \cdot C_k}{n} \quad (9)$$

where  $k=0,1,2$  standards for luminance and two chrominance channels,  $C_k$  is the dip-to color,  $n$  is the total number of frames in this effect, and the constant  $N$  is the DCT block size (default: 8).

This operation is directly applied to the DCT coefficients in I frames or DCT coefficients of residual in B and P frames. For a typical MPEG...  $I_0 B_1 B_2 P_3 B_4 B_5 \dots B_{n-1} B_n$ , with I/P frequency  $M=3$ , the operation for each type is:

$$I \text{ frame: } F'_k = F_k + i \cdot \Delta I_k \quad (10)$$

$$P \text{ frame: } F'_k = F_k + M \cdot \Delta I_k \quad (11)$$

$$B \text{ frame: } F'_k = F_k + \text{mod}(i, M-1) \cdot \Delta I_k \quad (12)$$

where  $F_k$ ,  $F'_k$  are the original and the modified DCT DC value, and  $i=0,1,\dots,n$  is the frame number.

#### Dissolve

The outgoing video fades out while the incoming video fades in. When there is no or low motion in the two videos, this effects can be approximated by the linear combination of the two video:

$$F(u, v, t) = \alpha(t) \cdot F_1(u, v, t_1) + (1 - \alpha(t)) \cdot F_2(u, v, t_2) \quad (13)$$

where  $\alpha(t)$  is a weighing function changing from 100% to 0%, user may modify it with any rate;  $F_1(u, v, t_1)$  is the last I frame of the outgoing video and  $F_2(u, v, t_2)$  is the first I frame of the incoming video. The resulting effect is a dissolve transition from a frozen frame of video 1 to another frozen frame of video 2. However, when either of the video contains high motion, re-encoding of few frames in the transitional period will be required.

### 5.2.2 Film Effects

Film effects refers to masking video with 4:3 aspect ratio to different aspect ratios such as 1:1.66, 1:1.85, 1:2.35, and 16:9. For I frames, the DCT blocks outside of the desired region are set to 0, and the blocks that lie on the masking boundaries are recalculated using the simplified DCT translation algorithm described in [7].

$$DCT(B) = DCT(H) \cdot DCT(A), \text{ where } H = \begin{bmatrix} 0 & 0 \\ 0 & I_h \end{bmatrix} \quad (14)$$

where A is an original block located on the boundary, B is the new masked block, and  $I_h$  is the identity matrix with size  $h \times h$ , as shown in Figure 7(c).

For P and B frames, only macroblocks with motion vectors pointing outside of the masking region need to be re-encoded. Macroblocks with motion vectors pointing inside do not need any modification. Efficient algorithms for re-encoding macroblocks are described in [7,8].

### 5.2.3 Key Effects

Key effects are often used for compositing an anchorperson with a scene, such as a weatherman in front of a satellite weather map. In spatial domain, this is done by shooting the first video with a uniform background color (usually blue), then replace every blue color pixel with the second video. In compressed domain, we segment the first video into foreground and background regions by detecting the blue color. Then we replace the macroblocks with just blue background color with corresponding macroblocks from the second video. We need to re-encode the macroblocks lying on the region boundary and the macroblocks with motion vector pointing outside their regions. The percentage of macroblocks which need re-encoding depends on the video type and MPEG encoder design. Some simulation results were reported in [7]. The complexity of the re-encoding process can be reduced by using the pre-existing motion vectors to infer new motion estimation parameters.

### 5.2.4 Motion Effects

Motion effects include *Freeze Frame*, *Variable Speed* and *Strobe Motion*.

#### Freeze Frame

Since the freeze effect is usually longer than 1 second, simply inserting duplicated frames (e.g. zero-energy P frames) for a long period of time is not desirable for interactive playback (e.g. random search) due to the lack of frequent I frames. We need to place an I frame at regular short interval. Therefore, the frozen frame is converted to an I frame if it were B/P frame. And the rest of the GOP is filled with duplicated P frames. All the macroblocks in the duplicated P frames are set to Motion Compensation Not Coded (i.e., 0 motion vector, and the 0 residue error blocks are not coded).

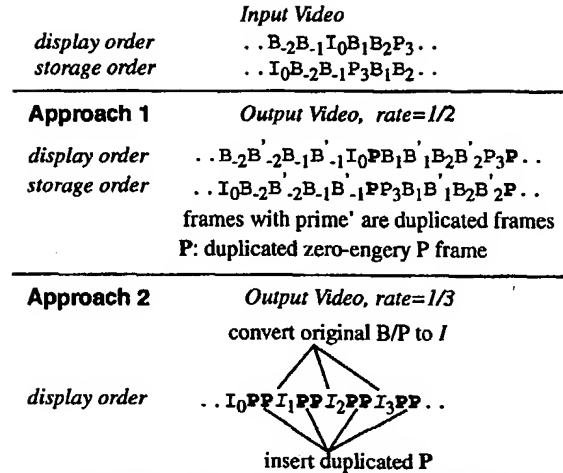


FIGURE 8. Two Approaches of Slow Motion Effect

#### Variable Speed

For fast motion, B, P, and I frames are subsequently dropped according to the variable speed.

For slow motion, depending on the slow motion rate, two approaches are used as shown in Figure 8. In approach 1, duplicated frames are inserted with no decoding involved. But the I/P frame delay is multiplied by the inverse of the motion rate. For example,  $I_0$  of output video must be transmitted 4 frames earlier, rather than the original 2 frames. This approach is suitable for rate 1/2 and up.

In approach 2, original P/B frames are converted to I frames using our DCT domain techniques [7]. Then duplicated P frames will be inserted between I frames. This approach reduces the frame delay, however extra DCT domain manipulations are required.

#### Strobe Motion

*Strobe motion* is a combination of *Freeze Frame* and *Variable Speed*. It is done by dropping original B/P frame and inserting duplicated P frames.

As described in Section 5.1.2, to avoid decoder buffer to overflow (e.g., inserted frame is too small) in constant bitrate video, we may stuff redundant bits to the inserted P frames. To avoid any buffer underflow, we may apply rate adjustment techniques described in Section 5.1.2.

### 5.3 Advantages of Compressed Domain Approaches

For the basic editing function: cut and paste, the compressed domain approach runs at least 60 times<sup>1</sup> faster than the straightforward approach (decode-edit-encode). That is based on 12 second per cut on average, one P or I frame for

1. Based on analytical estimation of computation complexity as well as software simulation results.

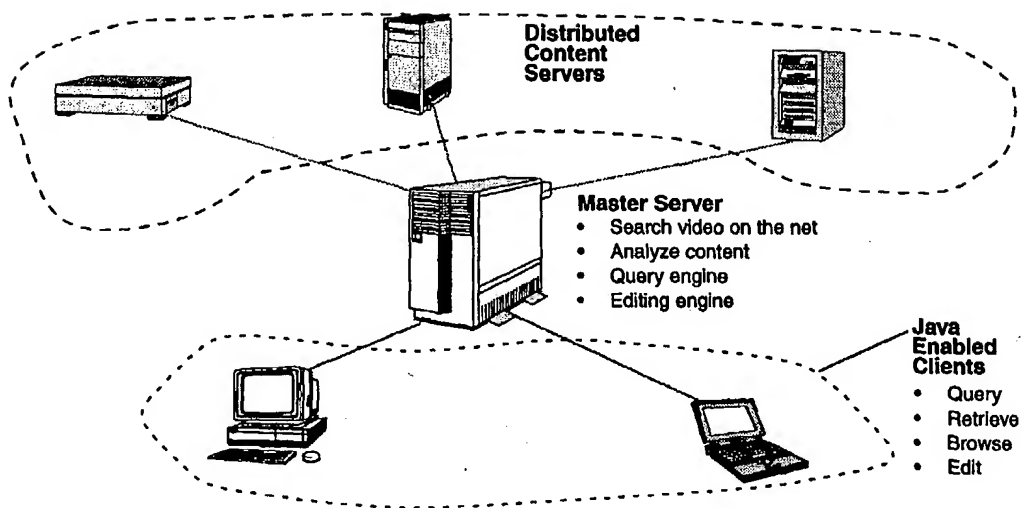


FIGURE 9. Client-Server Based CVEPS System

every two B frames, and cutting at arbitrary locations. The speedup can go over 600 times if we allow cuts only at P frames. The longer the segments are, the higher speedup we gain. The compressed domain approach also avoids quality degradation because the second quantization in the re-encoding process is avoided. For example, we observed an average 3.6 dB gain for a 60 frame segment (608x224, 4.0 Mbps). Only the re-encoded boundary GOP will suffer the 3 to 4 dB quality loss as in the straightforward approach.

## 6. SYSTEM DESIGN

The CVEPS uses a distributed client-server model as illustrated in Figure 9. The master server is linked with WebSEEK which searches for image and video files over the WWW. Once a video file is found on any other hosts or WWW distributed content servers, it will be downloaded and preprocessed by the master server to extract the keyframes and associated visual features such as camera motion, moving objects, color, texture, and temporal variance etc. The HTTP address of video and the extracted features will be stored on the master server. This client-server model gives the client much richer resources that are not constrained to the client's local environment.

The client is implemented with Java applets. The client may open any video at the server and browse the keyframes hierarchically using story structure or content clustering methods [26]. All the keyframes are hyperlinked to the WebSEEK's query engine so that the keyframes or objects may be used to form new visual queries for new videos or images over the entire master server.

To view the video, the user may simply drag the keyframe which represents for a video shot to the source monitor of the editing interface, see Figure 9. A low resolution copy of

the video shot will be sent to the client by the server. The client can use the interactive MPEG2 viewer/decoder to do random access, step forward, fast forward/reverse and normal playback. The MPEG2 decoder is written in C and compiled as a run-time shared library to be called by the Java client.

The user may also turn on the *VideoMap* option of the MPEG2 player. This option will invoke the display of the bounding boxes of any moving objects detected (described in Section 4.3). By clicking the mouse inside the bounding box, the client will send a request to the server to get additional information of the object (e.g. a hyperlinked home page) or invoke content-browsed visual query using this object as a template.

To edit the video, the user may mark in/out any segment of the video shot in the source monitor to splice-in or overwrite to the new sequence in the record monitor. A separate timeline window will show the resulting video/audio tracks and the detailed information of each included video shot. The user may also insert special effects as described in Section 5.2.

During the editing, only the Edit Decision List (EDL) is created. The new sequence must be rendered before it can be displayed. There are three levels of rendering. At the first level, the client uses C routines from its shared libraries to render only the straight cuts at low resolution without showing the special effect. At the second level, the client may send the EDL to the server for generating the new low resolution video with desired special effects. Finally, when the client is done with the editing, the master server will generate a full resolution video with all the effects from the highest quality source video which is located at either the master server or the distributed remote content servers.

## 7. CONCLUSION

We presented a Compressed Video Editing and Parsing System with our proposed compressed-domain video manipulation and indexing techniques. The CVEPS processes the compressed video to automatically extract key visual features such as scene cut, camera operation parameters, moving objects, and then visual features (e.g., color, motion speed and trajectory). Content based queries are formed with the above visual features for retrieving new video clips. The CVEPS also provides tools for editing compressed video and creating special effects. We have shown that the compressed domain approach can achieve significant system performance improvement in speed, quality, and storage. Software implementations of the proposed algorithms have been developed in C and Java employing a client-server model over the WWW. The client-server implementation is particularly useful for users with access to regular computers or even less powerful devices (such as light-weight mobile units).

## 8. ACKNOWLEDGMENT

This work is supported in part by a NSF CAREER award (IRI-9501266), HP, Intel, and the ADVENT project of Columbia University. Implementation of the JAVA editing interface of the CVEPS was contributed by Jan Stanger while he was a research intern at Columbia University. The video shot browsing and clustering interface was contributed by Di Zhong.

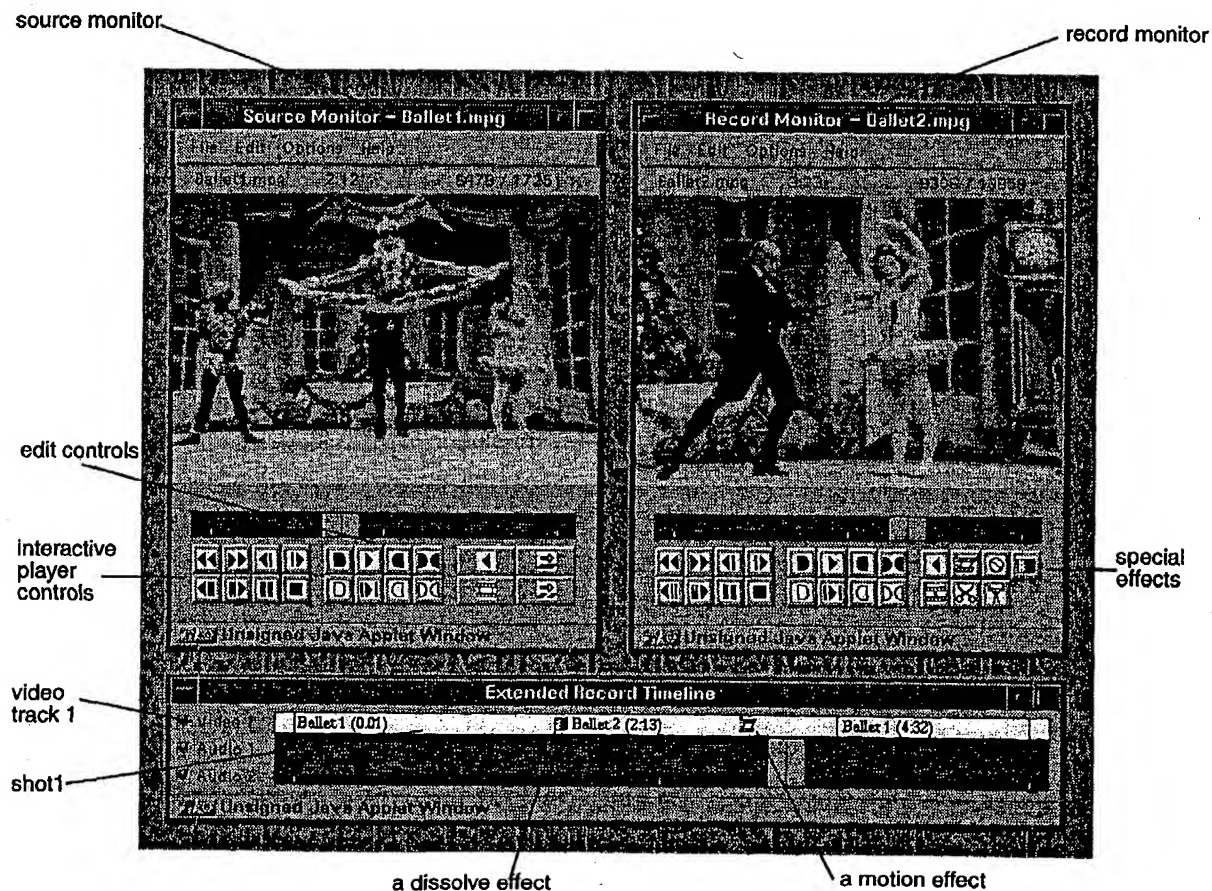


FIGURE 10. The CVEPS Video Editor Java Interface

## 9. REFERENCES

1. *AVID Effects Reference Guide*, Avid Media Composer and Film Composer, Release 5.50, June, 1995.
2. A. Akutsu, Y. Tonomura, H. Hashimoto, and Y. Ohba, "Video Indexing Using Motion Vectors," *SPIE Visual Communications and Image Processing 1992*, Vol. 1818, pp. 1522-1530.
3. A. M. Alattar, "Detecting and Compressing Dissolve Regions in Video Sequences with DVI Multimedia Image Compression Algorithm," *ISCAS 1993*, pp. 13-16.
4. F. Arman, A. Hsu, and M.-Y. Chiu, "Image Processing on Compressed Data for Large Video Databases," *Proceedings of ACM Multimedia '93*, June 1993, pp. 267-272.
5. J. Bergen, P. Anandan, K. Hanna, and R. Hingorani, "Hierarchical Model-Based Motion estimation," *2nd ECCV*, 1992, pp. 237-252.
6. S.-F. Chang, "Compressed-Domain Techniques for Image/Video Indexing and Manipulation," *IEEE Intern. Conf. on Image Processing, ICIP 95, Special Session on Digital Image/Video Libraries and Video-on-demand*, Oct. 1995, Washington DC.
7. S.-F. Chang, Compositing and Manipulation of Video Signals for Multimedia Network Video Services, Ph.D. Dissertation, U.C. Berkeley, Aug., 1993. Postscript files also available from <http://www.ctr.columbia.edu/~sfchang/thesis93>.
8. S.-F. Chang and D.G. Messerschmitt, "Manipulation and Compositing of MC-DCT Compressed Video," *IEEE Journal of Selected Areas in Communications, Special Issue on Intelligent Signal Processing*, pp. 1-11, Jan. 1995.
9. N. Dimitrova, and F. Golshani, "Motion Recovery for Video content Classification," *ACM Transactions on Information Systems*, Vol. 13, No. 4, October 1995, pp. 408-439.
10. A. Eleftheriadis, and D. Anastassiou, "Optimal Data Partitioning of MPEG-2 Coded Video," *Proceedings of 1st International Conference on Image Processing (ICIP-94)*, Austin, Texas, November 1994.
11. A. Hampapur, R. Jain, and T. E. Weymouth, "Feature Based Digital Video Indexing," *IFIP2.6 Visual Database Systems, III*, Switzerland, March, 95.
12. ISO/IEC 13818 - 2 Committee Draft (MPEG-2).
13. ISO/IEC/JTC1/SC29/WG11, MPEG Document AVC-400, Test Model 3.
14. J. Meng, Y. Juan, and S-F Chang, "Scene Change Detection in a MPEG Compressed Video Sequence," *IS&T/SPIE Symposium Proceedings*, Vol. 2419, Feb. 1995, San Jose, California.
15. J. Meng and S.-F. Chang, "Tools for Compressed-Domain Video Indexing and Editing," *SPIE Conference on Storage and Retrieval for Image and Video Database*, Vol. 2670, San Jose, California, Feb. 1996.
16. A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances," In E. Knuth and L. M. Wegner, editors, *Video Database Systems, II*, Elsevier Science Publishers B.V., North-Holland, 1992, pp. 113 - 127.
17. T. A. Ohanian, *Digital Nonlinear Editing: new approaches to editing film and video*, Focal Press, Boston, London, 1993.
18. H.S. Sawhney, S. Ayer, and M. Gorkani, "Model-Based 2D & 3D Dominant Motion Estimation for Mosaicking and Video Representation," *Proc. Fifth Int'l conf. Computer Vision*, Los Alamitos, CA., 1995, pp. 583-390.
19. B. Shahraray, "Scene Change Detection and Content-Based Sampling of Video Sequences," *SPIE Conf. Digital Image Compression: Algorithms and Technologies 1995*, Vol. 2419.
20. J. R. Smith and S.-F. Chang, "VisualSEEK: a Fully Automated Content-based Image Query System," *ACM Multimedia '96*, November, 1996.
21. J. R. Smith and S.-F. Chang, "Searching for Images and Videos on the World-Wide Web," Technical Report # 459-96-25, Center for Telecommunications Research, Columbia University, New York, August 1996. Also submitted to *ACM Multimedia Magazine*.
22. S.W. Smoliar, and H.J. Zhang, "Content-Based Video Indexing and Retrieval," *IEEE Multimedia*, summer 1994, pp. 62-72.
23. J. Swartz, and B.C. Smith, "A Resolution Independent Video Language," *Proceedings of ACM Multimedia '95*, pp. 179-188.
24. M.M. Yeung, B.-L. Yeo, W. Wolf, and Bede Liu, "Video Browsing using Clustering and Scene Transitions on Compressed Sequences," *IS&T/SPIE Symposium Proceedings*, Feb. 1995, San Jose, California. Vol. 2417, pp. 399-413.
25. Y.T. Tse, and R. L. Baker, "Global Zoom/Pan Estimation and Compensation For Video Compression" *Proceedings of ICASSP 1991*, pp.2725-2728.
26. D. Zhong, H.J. Zhang, and S.-F. Chang, "Clustering Methods for Video Browsing and Annotation," *Storage and Retrieval for Still Image and Video Databases IV, IS&T/SPIE's Electronic Imaging: Science & Technology 96*, Vol. 2670, San Jose, CA., Feb. 1996.



## Panel Session

## Future Directions in Desktop Video

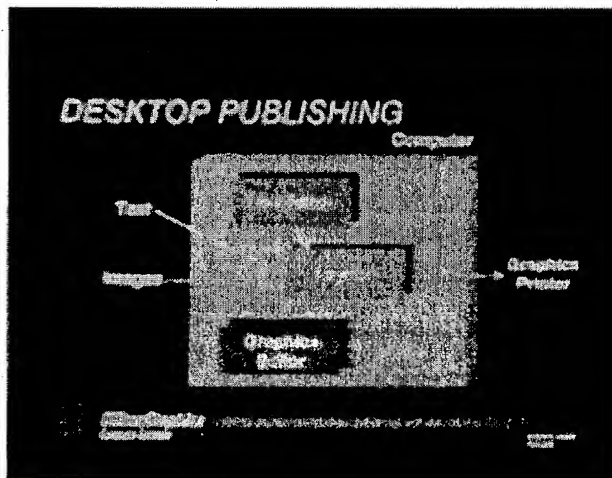
Chair: Tim Heidmann, *Silicon Graphics*

Speakers: Michael MacKay, *Sony*  
 Gregory MacNicol, *Computer Graphics World*  
 Floyd Wray, *BYTE-by-BYTE*

Good morning. My name is Tim Heidmann and I'd like to welcome you all to this panel, which is entitled Future Directions in Desktop Video, and I'd especially like to thank all you people who stayed up a little late on Thursday night to come to this panel. It's really good to see you all out there.

I've gotten word that this panel is being transcribed. They're putting together a booklet, so they're taking the slides and the stills from the videos and all the things that we're saying. So I'd just like to take this opportunity to say hi to the person who's transcribing this and sorry you couldn't be here today, and I wanted to let you know that the word of the day is Neopraseodymium, and I hope you've got your scientific dictionary close by.

When we first started putting this panel together, I talked to my friends who were involved in a number of different areas in video, and the question that came to the forefront very quickly is what exactly desktop video is. There's been a lot of talk about it, a lot of magazine articles. It's a good buzz word. But we all felt it incorporated a whole bunch of different areas that weren't easily put into one category.



— HEIDMANN - SLIDE 1 —

### DESKTOP PUBLISHING

Incorporates elements from outside world

Entire product can be created within computer

Utilizes computer software like text editors and graphics design

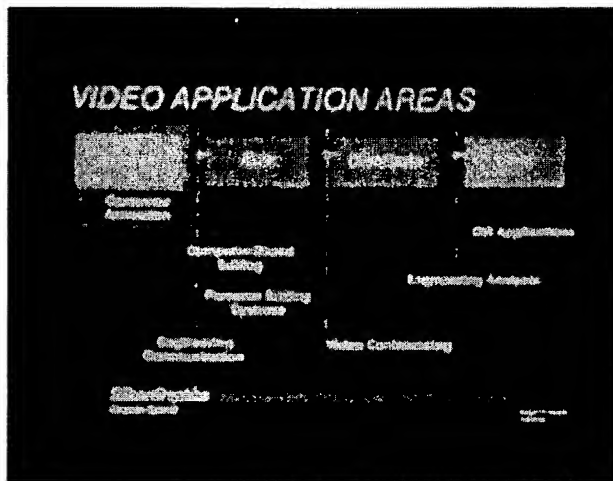
Made possible by high quality graphics printer

— HEIDMANN - SLIDE 2 —

We did agree that the name of desktop video came from the field of desktop publishing. In desktop publishing, which has been a rapidly growing field in the past few years, the whole point is that we've got a computer bringing together elements from the outside world, creating elements inside the computer, putting them all together and coming up with a final product. The point is it's all done inside the computer. Again, it does the things that computers do really well -- like text editing and graphics design and layout. And it was made possible by the fact that these high quality printers -- laser printers -- had come out that you could produce a very high quality output from it.

Well, on the video side, there is a similar development. That is, it's possible now to make video animation completely within the computer. There are software packages for modeling objects, for creating animation, for rendering very high quality images and outputting them directly to tape. And I guess you could call that desktop video. You're doing the same thing as you're doing in desktop publishing, but now you're producing videotape and animation.





— HEIDMANN - SLIDE 3 —

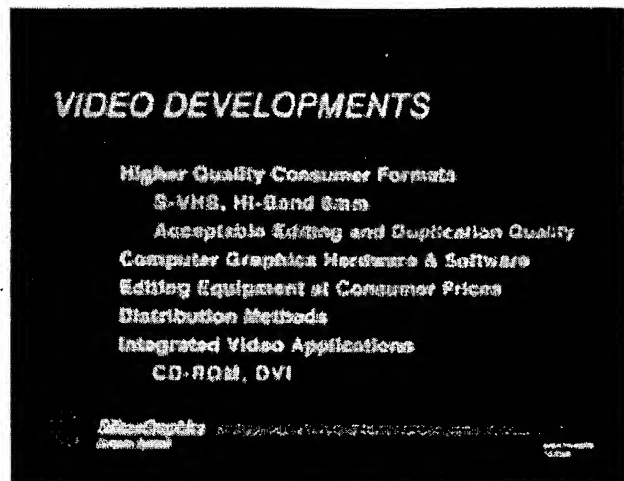
But really what's happening in video is a lot bigger than that. I've kind of come up with this map. If you look at the entire video process, you can split it into four parts. The first being creating the elements, which I've called Source here. Now that would include such applications as computer graphics, generated completely inside the computer, but also things like pointing a camera at someone or something, things like medical imaging. Basically the creation of the images.

The second step would be assembling those images and probably some audio into a master video production. Just about everything you do in video involves some sort editing to it, even if it's just putting a title on the beginning.

The third area is the distribution. How do these images get out to the outside world? A lot of times it's just making lots of VHS copies and mailing them to your friends.

And the final part is how do you look at this videotape? How do you use video in your application?

What we're going to talk about today, when we talk about desktop video, is actually all these areas. The speakers today have a number of different backgrounds. We'll be addressing this in a number of different ways.



— HEIDMANN - SLIDE 4 —

Basically the reason we're doing this panel today and why it's important now is because there are a lot of developments that are bringing video into the reach of more application areas. People are interested in what can be done with video, want to know what's happening and what the developments are. Specifically the things that we're seeing are the appearance of higher quality consumer formats. That is, videotape recorders, players, that are available at consumer affordable prices, but give you enough quality to let you duplicate and edit a little better than VHS or just plain 8 millimeter.

Computer graphics hardware and software is becoming less expensive and more accessible. The fact that you can buy editing equipment for these new consumer formats and do a really good job of putting together a final videotape without going to a postproduction house. New distribution methods and integrated video applications like DVI, which allow much broader use of video. So there's more incentive to produce these things.

These are the areas we're going to be talking about today, and just keep in mind this map, and I think if we can reference the different things we'll be talking about to this map, maybe it will all make sense.

I'd like to introduce the first speaker today who is Michael MacKay, currently of the Sony Advanced Video Technology Center, formerly of Diaquest; before that with Atari Research. Michael brings a strong background in computer graphics, in video production, interactive video applications, and without further adieu, Michael.

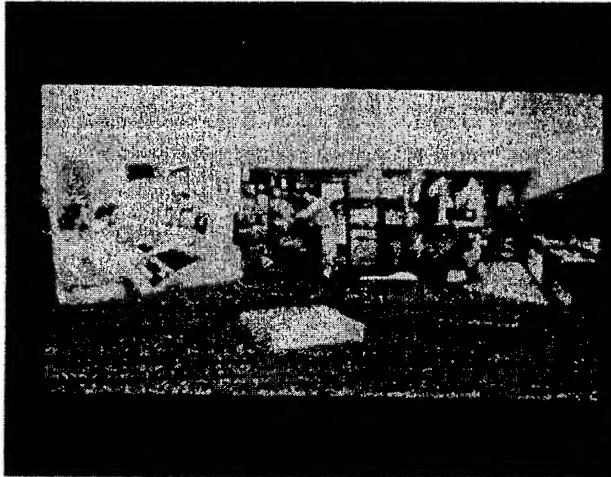
**Michael MacKay**  
Sony

Good morning. What I'd like to do first is to get Tim's slide pulled back up that showed the map. Maybe on that screen if you can do that. I don't know if that's possible. I've got one tray.

What I did here is -- that's fine. Back that one back up. That's great.

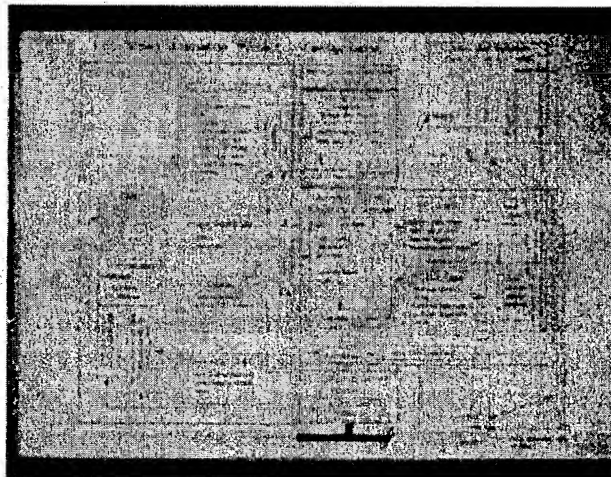
When looking at desktop video, there's a range of applications and a range of things that one could consider dealing with, and one of the first things that I thought of in a desktop video environment is something that allows the user

to empower himself to command his own environment. Typically you go into these production environments or postproduction environments and there's just so many people involved, it becomes more of a political issue to get along with all the people involved to get the project done, as much as it is to deal with the content and elements involved. So I put a little slide together here on the first one. I'm going to have my tray come up on the righthand side here if I can, and go forward.



— MacKAY - SLIDE 1 —

Really what you want in the desktop video environment is this. So I took a little desktop and this is the one-inch facility that we built at Atari Research, and this is the kind of resources that you really desire to produce the kind of interactive programs, as well as the linear programs that you would use in an environment. So this is a little bit of a spoof kind of outlining the kind of capabilities that a person would like to command from their desktop.



— MacKAY - SLIDE 2 —

When you get into what a facility like this looks like, this isn't intended to be readable, but it shows you all the different peripherals, all the different kind of audio devices, video devices, signal processing devices and items, that one needs to

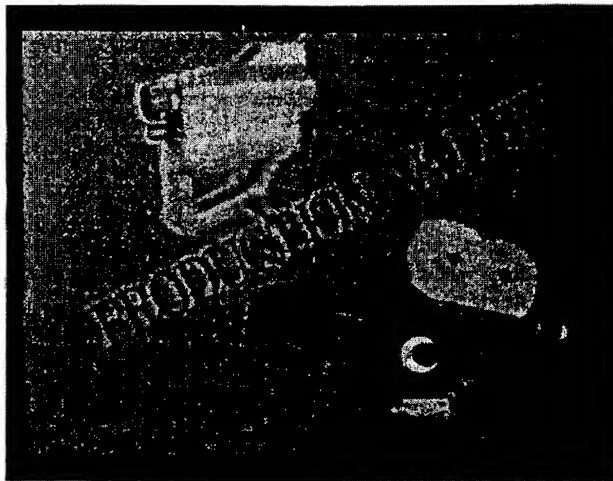
have control of to put together a basic multi-media production. When you add in layers of interactivity, it adds a whole new layer of complexity of visual and audio data management.

I did some work with a gentleman at Lawrence Berkeley Laboratories, named Bill Johnston, which represents a typical application of a desktop video environment. If you look at the bottom of the slide there's a little thing that says "To Cray", and he literally was doing simulations and calculations on a Cray computer, and those simulations were being downloaded over Ethernet to a VAX that was then transferred to an IBM PC. I have a demo reel that shows some of the animations that he was able to generate. The ironic thing of this situation was a Cray was feeding images to a Sony Betamax deck. So all of this high powered hardware on one end and his objective was to have the lowest cost transportable medium so that he could go home and watch these simulations. And they were simulating fluid dynamics and other very complex relationships. To see a Betamax tape deck connected to a Cray was quite an interesting application. Considering they had significant resources and could have gone with any video format they wanted. The objective here was to get it down in the hands where he could take the thing home. Mr. Johnston wanted to be able to go into a boardroom meeting or a meeting with his colleagues and not have to pull them over to the terminal that was hooked up to the super computer.

The next slide shows a generic desktop video environment. Most computer systems do not generate line rates and signals that are compatible with videotape recorders, anybody that's created a full color picture of any resolution quickly sees how much data these things take and unless you have stock in a hard disk manufacturer, you probably don't want to keep all these images on your hard disks. There are things happening with mass storage devices that will allow you to store this images, but still nothing is capable of 30 frames per second real time playback. So we still like to control videotape recorders as a -- what I call a video peripheral now -- that is another storage device that hangs off the computer. And there is some transcoding or encoding or some kind of signal mucking you have to do to get over there, and this allows you to do that under computer control -- control this environment.

The nice thing about these environments now is that you don't have to be there. You don't have to be sitting there pushing the buttons and doing everything. You can set up batch files or you have hooks in the application programs where basically once you set up the environment, describe what kind of imagery you want to generate, the computing platforms will then go through and push all the appropriate data out to the videotape recorder. And the nice thing is when the videotape recorder rewinds and cues, you can actually see if it's making it there.

So it gives you a nice little feedback loop and you can walk home for the night and you come in in the morning and hopefully you're pleasantly surprised; sometimes you're not. And you have to rerender and go over again.



— MacKAY - SLIDE 3 —

When you're entering into these desktop video kind of applications, there is a real question you have to ask yourself of what your expectations are. There is this question of production value, and what I've done here is generated a slide in which the camera up at the upper left here is actually a Fisher Price video camera called Kiddie Video, and it's black and white recorded on a Phillips cassette. Some productions may not need more than Kiddie Video. On the other end, there's the high end Panavision cameras, and so you need to make some decisions of where this thing is going to finally get distributed. Am I going to be able to do the final product in my desktop environment, or am I going to use this as an offline tool that allows me to prepare, make some decisions in my personal environment. But then I will go out and use the high end, the high quality material, to go off and generate my final program.



— MacKAY - SLIDE 4 —

Then that goes back to the question of whether you're using the system in an online or offline environment. These terms are loosely used these days because I can tie up a

complete one-inch facility at the biggest houses in the world and use it in our offline environment.

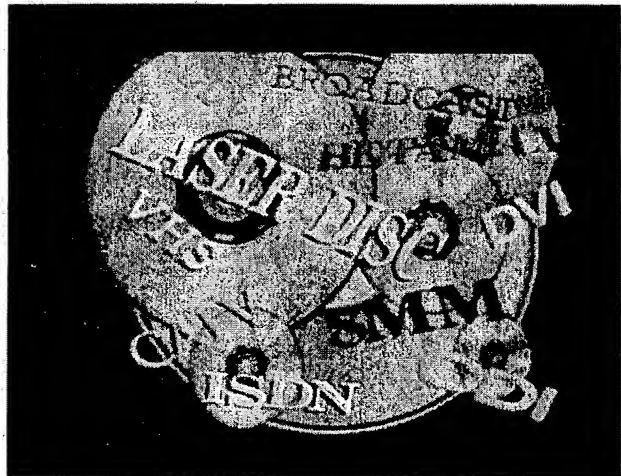
Online means that you're going for the final product and you basically tie up as many resources as you can to maintain and preserve the image quality. So you do not want to go multiple generations, especially in the analog domain, because this degrades the picture quality.

So what you try to do is you try to start in an offline environment, which is becoming -- many tools are becoming available. There is many products being made that reside on the Macintosh platforms and SGI and Ataris, Apples, PCs -- all sorts of stuff that you can do.

These tools allow you to basically look at image data, correlate that against the time code number, which is how you address the frames, and then what you can do is build yourself an edit decision list. Then you go into the online environment.

The main drawback in doing this right now is many of the kind of elements that you'd tie together in an online environment are not documentable in the offline environment so that it's machine readable. In other words, you have a lot of comments about special effects, a lot of comments about keying, and these kind of layering effects.

So one of the things that the industry is looking at right now is going to a next generation of edit decision lists that would allow you to literally store every parameter of every device in the entire room, and recreate the situations. What that's going to do for everyone here is allow their desktop video environments to become much more powerful and much more meaningful in communicating to the online high quality environments, as well as improving the effectiveness of an offline environment.

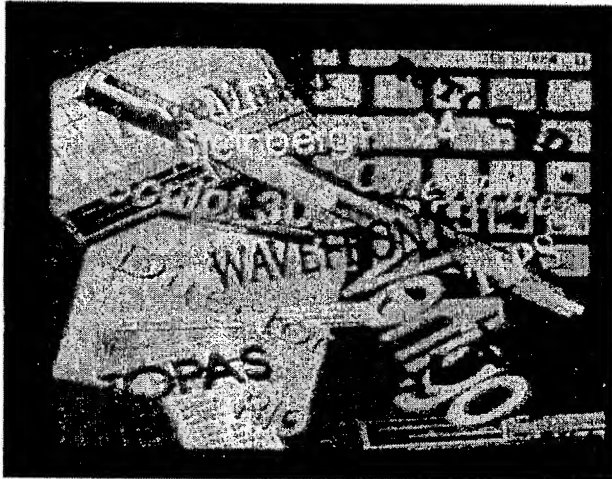


— MacKAY - SLIDE 5 —

One of the things I've been involved with for so long is that we were always trying to generate a new delivery medium, and this is a montage of all the different delivery mediums-- even the phone companies now are coming out with broad band ISDN, and they're hoping to be able to shove video images down the phone line in the not-so-distant future. There's been so much work in data compressions and DVI kind of strategies and standard distribution formats that there's plenty of tools for us to now disseminate this kind of visual information. But in my experience in this industry, the real drawback is to be able

to get an environment together where you can cost effectively produce the high quality program that you could distribute to these vehicles.

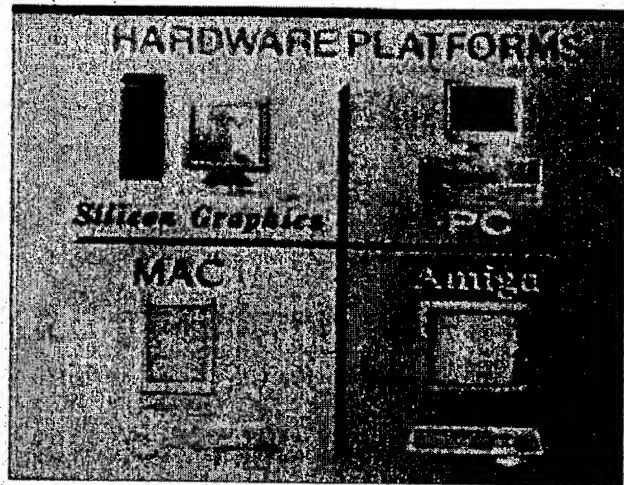
I've done a lot of work over 10 years in laser disc -- producing interactive laser discs -- I've yet to produce a laser disc that was under \$100,000. So this limits the amount of projects that people can do. I think that as desktop video comes of age when the tools really become available in the hardware and software categories. It will empower people like yourselves to take an assertive stance within your company and actually produce something that is not a compromise in the quality. You've been able to offline and offload a lot of the high dollar decisions because you can make those decisions in a more cost effective environment, and I think that will really fuel the ability to feed all these distribution networks.



— MacKAY - SLIDE 6 —

There's a wealth of different software programs that are just coming out. This slide gives a glance at some of these software tools. I'm not trying to push anybody's hardware or software here. But there's no reason to wait. I mean, there's a lot of things you can do right now. There's a lot of hardware available right now. There's a lot of software available right now. And if anybody wants to see me after or there's recommendations, everyone here on the panel has their own preferences but there are a lot of options and they can offer a lot of choices. So it's not something that you're going to have to wait another five years until it's practical.

In fact, all the slides that we've produced here utilized these graphic tools. It's a really nice environment where you can do the desktop publishing, video material, and then you want to print the label for that videocassette tape, or the label for that disc. You can use the same image data, the same thing, and it becomes a really productive environment.



— MacKAY - SLIDE 7 —

I just put up a couple hardware platforms here that I have personally used to create video programs, and each of these environments has their own plusses and minusses and I particularly don't believe in the position that most people take is to subscribe to one and one and only platform. I'm more into the distributed architecture thinking, where you want to eventually migrate. You can get in very inexpensively and migrate through a whole process of equipment and peripherals.

Sony's producing a bunch of machines that are now computer controllable under RS422. There are a bunch of formats. These are widely used in professional video, but also these tape machines are now under \$10,000. And in a corporate communications environment you get very good control of these tape machines under computer control and can perform a variety of functions.



— MacKAY - SLIDE 8 —

One of the new things, as Tim mentioned before, is the emergence of a high band 8 millimeter, and the nice thing about this format is that you can go out and shoot with a standard consumer camera, you can bring that tape into this environment, and you can poststripe the time code, frame

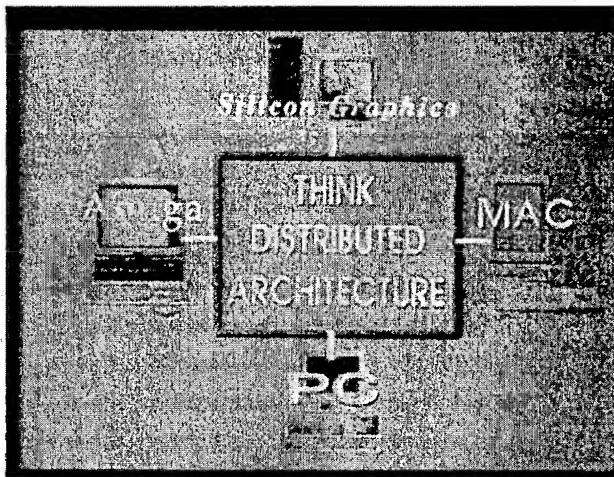


accurately, after the fact. So you don't have to run time code equipment in the field and have the expensive equipment, and this deck is also computer controllable by a computer.



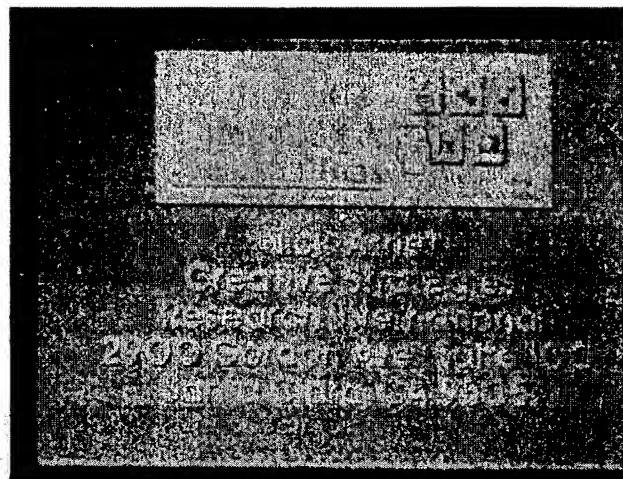
— MacKAY - SLIDE 9 —

Automatic editing equipment is available, switchers. These are all under RS232 control. It can be hooked up to any computer. In the higher end environments there are now totally computer controllable VTRs that allow you to integrate in the high end environment, and eventually what you'll do is you'll migrate your data and your images to these machines for certain distribution things.

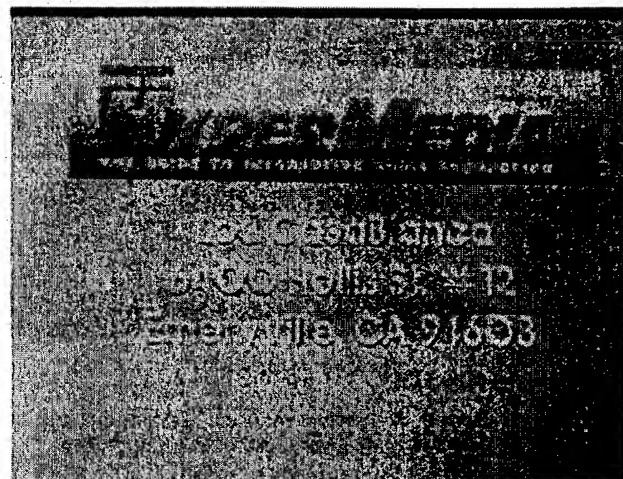


— MacKAY - SLIDE 10 —

So Silicon Graphics and Macintoshes, PCs, Amigas, NeXTs, whatever you have -- each is going to have special strengths, and I think the thing to do is to think about getting these things all connected together and use each machine for what it's best suited for.

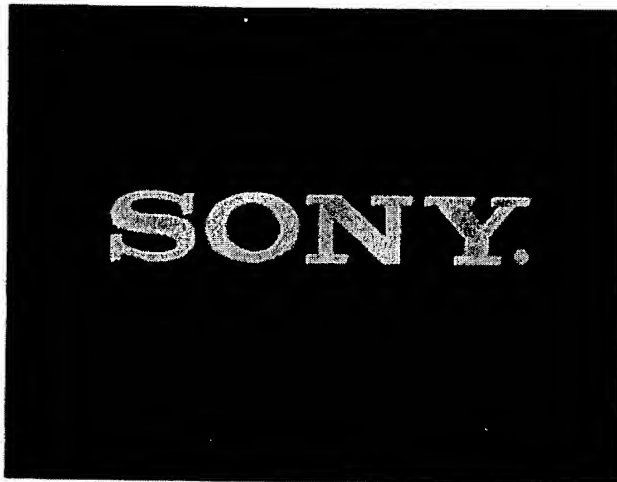


— MacKAY - SLIDE 11 —



— MacKAY - SLIDE 12 —

There's a couple places that I can direct you that I have found very helpful for getting other information, and one of these is the Multimedia Computing and Presentations Newsletter by Nick Arnett, and also I've done a lot of work with Lou Casibianca on producing this Hypermedia Magazine -- trying to disseminate more of this information, saying that the time is now that you can do stuff with this stuff right away, and in fact there is a large article that I wrote for the next issue of Hypermedia that basically goes through a product selector kind of guide to show you what kind of products are available in both display adapters and other kinds of equipment that you can use in this kind of thing.



— MacKAY - SLIDE 13 —

We're going to queue up a videotape here we've only got a couple minutes left, but we'll quickly show you some of the applications that I've been involved with that have used basically desktop video platforms and everything you're seeing on this tape was produced on a PC compatible platform. So if you could roll the tape, please.

— VIDEOTAPE BEING PLAYED —

This is the piece done by Lawrence Berkeley Laboratories from the Cray. This is from the University of California in Berkeley, which is sharing the videotape machine over an Ethernet network.

We're going to move on to the next speaker now. Thank you very much.

**Moderator**  
**Tim Heidmann**  
**Silicon Graphics**

Thank you, Michael. I was hoping to have time to ask a couple of questions in between speakers on that particular talk, but I think we'll wait until the end and get a chance to have a discussion after everyone's had a chance to speak. I think it will be a little simpler that way.

The next speaker I'd like to introduce is Gregory MacNicol. Gregory is an independent consultant and writer. His articles have appeared this year in magazines like Computer Graphics World, where he did a cover story on desktop video -- strangely enough. Also a consultant on Video Installations and Video and Computers. Gregory.

**Gregory MacNicol**  
**Computer Graphics World**

What I'd like to do is before I get into any technical stuff, why don't we roll my first tape.

— VIDEO TAPE BEING PLAYED —

Thanks. So that's the future of desktop video. As a writer, I am forced to describe and define in great detail what desktop

video is and what it's going to be. This is difficult because when I talk to various people, for every 10 people I ask, I get at least 12 different answers.

Desktop video is unfortunate because the term doesn't describe really the full meaning. For instance, when people do a simple animation like this, though this particular animation was done on an IBM PC, very often people forget the technology behind it. In other words, it's not just the computer -- it's the video technology.

A simple example is when I asked various people what did it take to create a certain video, and I describe in the article what computer equipment was used. It turns out that the video equipment was by far a lot more expensive. And this is one thing that's often overlooked and is a very key problem with what desktop video is right now. Can I see the first slide?

This is one example of a typical -- well, maybe not a typical -- configuration, but a configuration of a computer system integrating into a video output. If you segregate what a computer can do, and its capability of creating a high resolution graphics image, that's fine. But now with a video system you have to deal with synchronization and all the parameters of interfacing with the true NTSC signal, as well as a videotape recorder. Here's an example.

Say somebody created a nice video and everything looks fine -- it looks just great. They have created various segments and now they have them on three-quarter inch videotape, and then they want to edit that. They bring it to a postproduction facility and in spite of the signal looking perfect, the postproduction facility may recognize that the video signals don't match. In other words, when edited none of the colors fit. None of the colors work together. Furthermore, the edit in and out points don't fit. In other words, it comes down to the video. While it looks perfect from the computer standpoint, it is completely unusable for any editing. This is an incredibly serious problem for someone working on let's say an Amiga system or a Mac system or another PC system.

For that reason -- for me -- I look at desktop video as being basically two ways of looking at it. One is single frame animation and the other is sequential animation, which you can see in real time. To give you an example of that, can you roll tape number two please?

This system on tape number two is done on an IBM PC using Autodesk's Autoanimator, and will give you an idea of what it takes to do something in real time.

— VIDEO TAPE PLAYING WITH MUSIC —

Again, this is being run in real time from an IBM PC on a VGA board.

— VIDEO TAPE PLAYING WITH MUSIC —

Other than putting the pieces together of these different animations, there was no other editing used. Okay, you can cut.

That particular system was introduced here at SIGGRAPH this year. Those people who were very interested in doing desktop animation, I really want to warn you of the awful difficulties with working with video. I write about it and I describe a lot of the technical details. In fact, Michael MacKay has been very involved with the arduous, in fact, painful experiences that so many systems -- so many people have had in using high end systems. Could we see slide number two?

This is a real simple example of where you can see an IBM PC system being used to create reflection and refraction and the

kinds of things you would typically see on broadcast T.V. Here is an example of where the computer is certainly apt at creating a very nice image. But turning this into video is a lot more difficult. For instance, in even an industrial application when people say we would like high quality -- well, what do you mean by that? Well, we need this to be one inch. They don't know either what it takes to create one-inch quality, and very often the client will encounter equal awful problems simply because the client and very often the developer of the animation doesn't completely know the awful complexities with video.

This image here is done with the Amiga, and it indicates again reasonably high quality -- in spite of its low resolution. And some wonderful films have been done on the Amiga but they too have had the awful experience of being transferred into medium grade video.

Various people asked me -- well, how did you feel about the film show -- and the thing that was so striking to me was that there were so many technical flaws in so many of the films.

For instance, one film from IBM, for instance, their film had -- if you look at a straight line -- the film had great difficulty in creating a straight line. The film was always jiggling. For any company to produce a high quality video and omit the obvious -- in other words, creating that high quality video signal -- is almost inexcusable.

This is an example of where again using a simple PC system, something that's going to be a lot more common in desktop video in the coming years -- mixing live video with computer-generated objects.

This image here is from the Digital Art System and I don't know if you have all seen the film show, but this -- rather the TrueVision Show -- they showed this and it's reasonably advanced. They're showing the full level of character animation and incredible detail too. It's a very impressive film.

But in order to get this kind of quality, in spite of the system costing -- let's say of the Amiga system, which is the least expensive system -- you still need in the area of several thousands of dollars of video gear.

So the future of desktop video is really going to be relying on two aspects. One, interfacing the computer with adequate video equipment -- video equipment that will accept still frames without synchronization problems.

Secondly, what we're going to be seeing is a pretty impressive aspect that's really going to open up a lot of capability, and that is the omission of very expensive video hardware. For instance, with 2-D animation systems such as the Auto Animator, very complex animations are going to be possible without needing still frame animation capability. In other words, connecting your computer directly to the video hardware so it can be used on the least expensive systems, such as 8 millimeter and such.

So as far as the future of desktop video is concerned, the problem that I face and I think the problem that all of us are going to be facing is defining what it is. Thanks.

**Moderator  
Tim Heidmann  
Silicon Graphics**

Thank you, Gregory. I'd like to say a few words now about a couple of application areas. I work for Silicon Graphics in marketing, and I deal with our customers who are doing very high quality computer-generated imagery -- usually for broadcast, entertainment and commercials -- that sort of thing.

So I'd like to talk about first of all what's happening in those fields and how it relates to what's going on in the PC world. And secondly, the directions we're trying to go in scientific and engineering use of high quality video.

## COMPUTER IMAGERY APPLICATIONS

Entertainment  
Presentation  
Communication  
Training

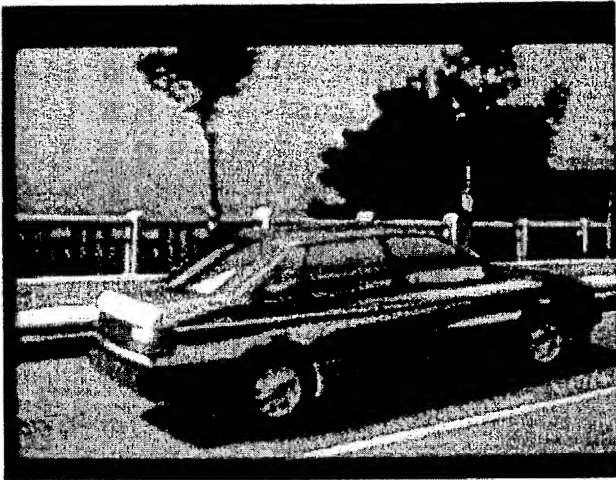
— HEIDMANN - SLIDE 5 —

Are my slides up? Thank you. You're all familiar with the entertainment applications of high quality video. We've seen a lot of it in the film show. Specifically the areas that we're trying to move into are presentation, communication and training. What does that mean? Well, presentation -- and the reason I split them up like this -- presentation tends to be very high quality animation. It's the sort of thing you would show to a customer in order to get him to spend several million dollars. This is a use of video used by very large companies like Boeing and McDonnell Douglas. When they're selling a fleet of planes they will do an animation of the plane with the company's logo on them and get them all excited so that they're eager to go ahead with the deal.

Communication, on the other hand, is the use of video to work within a large company -- say, very large engineering teams. Building a satellite, for instance. To let everyone know exactly what's going on, what it is they're trying to build. The emphasis here is more on interactive use of video, rather than on the very high quality image producing of videotape that you use over and over and over again. For communication, the quality of the final product is not quite as important.

And finally, training. Again, you can get by without a lot lower quality final product and the difficulty here is relating it to the real process and making a very realistic looking videotape.

On this side I've just got a few slides of various industrial applications.



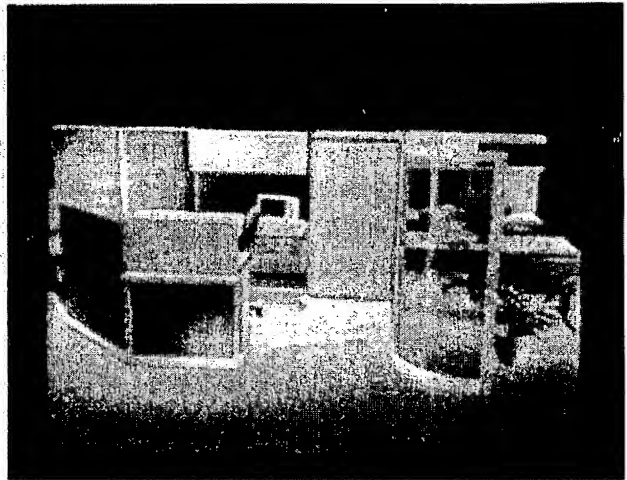
— HEIDMANN - SLIDE 6 —

The first was auto design.



— HEIDMANN - SLIDE 7 —

Here is architectural previewing.



— HEIDMANN - SLIDE 8 —

Interior office design. To give a couple of credits, the first slide was from Alias Research Software. These past two are from Thomson Digital Image.

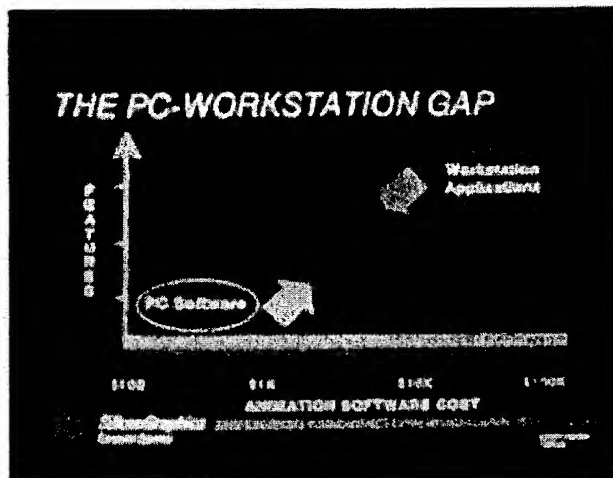


— HEIDMANN - SLIDE 9 —

Another Alias slide for consumer design.

There is a tradeoff in what you can get versus what you pay for. Those images you saw were very high quality -- almost photo real in some cases. They also cost a lot of money. They're probably done on workstations that cost \$75,000-\$80,000 with software that costs maybe \$40,000-\$50,000. It requires a fairly trained technical person dedicated to using the software. That is indeed an expense, and that's a problem.





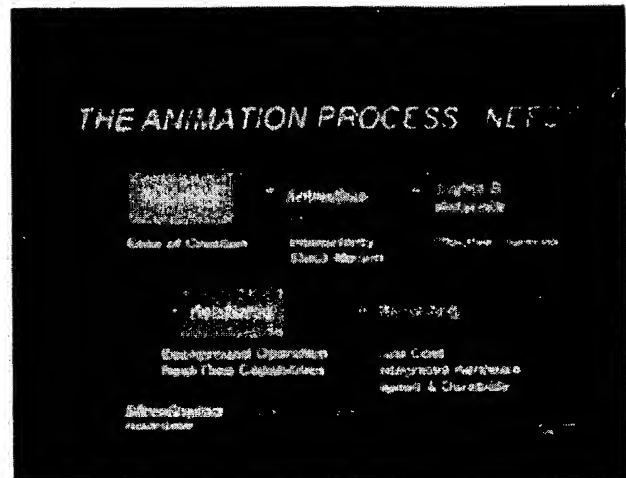
— HEIDMANN - SLIDE 10 —

There is a gap in what you can buy today. PC software that generally fits in the \$100 to \$1,000 range -- software is a much smaller percentage of what you're paying for the computer hardware. Certainly if you're including the video hardware, the software is a very small percentage.

On the other hand, the workstation applications like Wavefront, Alias, Vertigo -- very high dollar applications riding on high dollar workstations. And there's nothing really in between. There are movements in both directions but there's still quite a gap there. I think the real majority of applications that we'll see developing in the next three to five years are going to be in that space in the middle, and who's going to get there first?

A couple of things that are happening is some of the very high end software companies are starting to develop lower cost packages sort of in the \$5,000 to \$10,000 range geared at industrial design and scientific markets. You can go out to Wavefront now and buy a package that costs I guess about \$5,000 and do very high quality rendering. It's still designed to run on a workstation, but it doesn't cost the \$10,000, \$15,000, or \$20,000 that you'd need to spend to buy their broadcast quality renderer. Let's see here. Okay, we don't need that one.

I'd like to talk just a little bit about the animation process and the things that I think ought to happen to make it easier. So if there are any software developers out there, please work on this. And hardware developers too. There are definitely some hardware concerns here.



— HEIDMANN - SLIDE 11 —

The animation process -- basically modeling, animation, designing your lights and materials and applying them to the objects, rendering your final frames, and finally recording it onto videotape.

Modeling, or for that matter the import of objects from the outside world or the import of images from the outside world -- is just a really tough process. There are modelers that are very easy to use and can produce a fairly limited set of objects. I view it as a toolbox that there is no one modeler that will do everything, but you've got to have access to a lot of very easy-to-use tools, a lot of different shapes, and things like -- think about modeling the human body. How would you go about doing that? I haven't really seen a modeling package that will let you get a good looking body unless you sculpt it out of clay, and digitize it in three dimensions.

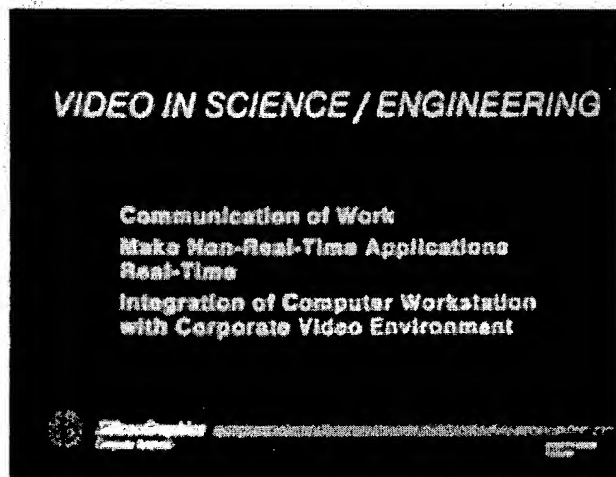
Animation is really tough to do, especially if it's not an interactive process. Animation fundamentally has to be interactive if you want control and expression in the motion that you're trying to get. So the first thing is interactivity. What kind of hardware can we put into systems to be able to play back 3-D animation at speed, to be able to change motion descriptions while you're watching the motion, to be able to interact with animation curves if that's what you're trying to do. The ability to animate images. Basically a digital video effects box inside a computer would be nice. The availability of stock motion tools in animation. That is, how many times have you seen a logo come in and seat like that? Well, it's tough to do if you do that from scratch. You really should have a button to do that.

In picking lights and materials, that should all be done for you and you should -- everyone kind of has on their shelf their favorite stock green marble and brown wood to be able to pull those off and just click on it and there it is.

In rendering the important thing about rendering on the high end is the availability to farm it out to a lot of different machines, to have it run in the background, to have complete control so you don't worry about it so you can animate and have this stuff going on. You can go home at night, it's going on. The availability to do that really easily. If you're going to do complex rendering, that should be easy. Real time capabilities. Gregory mentioned this and I believe it's of the absolute importance in that middle range. Most of the video

that's going to be produced from computers is going to end up being what you can do in real time, because frame by frame, single frame recording is so hard and it takes so much time and it ties up your machine for hours. If you can do it in real time, you save a lot of trouble and computer animation becomes a short turnaround tool that you can really use in a corporate environment. And finally, recording -- low cost. Right now you have to put the components together yourself and that tends to make each little piece cost a lot of money -- encoders, animation controllers, the deck itself, monitors. Turnkey solutions are going to be very important. I think computer companies need to realize that people are trying to use their systems to do this stuff. They need to take the responsibility to put it together. Computer companies -- maybe it's the video company that should do it. I think Sony should make a deck that you should just be able to plug right into a computer. Video port, RS232 port, so that you can do single frame recording. (Applause) I hope there's someone from Sony here. Oh, Michael, you're here. I know Michael already feels this way, so I'm just selling a sold person right now.

Speed and durability. When you start using video equipment with computers you beat the heck out of it. We thought editing was bad. Using it on computers is really tough, and when you've already got your hands full with trying to keep your computer running, the last thing you need is to try and figure out what's going on with your video deck.

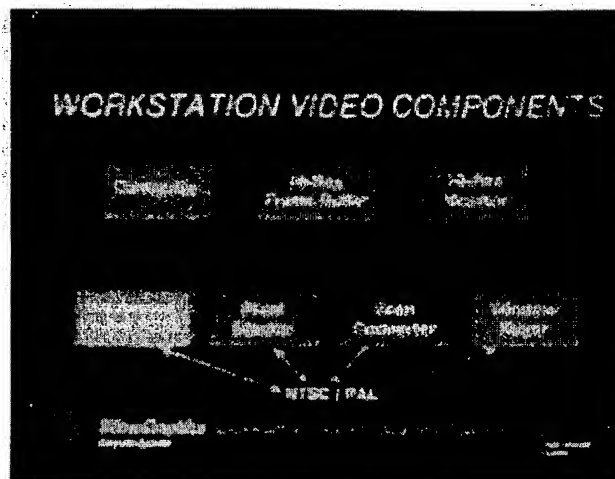


— HEIDMANN - SLIDE 12 —

Let's see what's over here. The second half. Video in science and engineering. Why would you use video in science and engineering? You've got this \$200,000 computer. Why would you need a video deck? Three major reasons that we see. First of all, so that you can take the work that you're doing, bring it to a conference, show it to your colleagues -- here's this program I'm running, you can see this little result. No big deal; I've got it on my screen. I just want to take it and show it to somebody.

Secondly, to make non-realtime applications realtime. No matter how much compute power you've got, you're going to be able to come up with a simulation that takes five or ten seconds a frame -- several minutes a frame -- to run and produce the graphics. You'd like to have some kind of machine on your desk that can let it run this simulation, record all those frames up. Then you can just sit there and look at them forward and

back, and be able to pick out the details -- whatever it is you're looking for. Thirdly, integrating what you're doing on your computer with what's going on in the rest of the company. With video, cameras. If you're doing simulation, robotics, control, visual simulation -- all that usually involves video. You'd like to be able to have video on your computer screen, to be able to combine graphics with video, to be able to send graphics out to various video sources -- distribute it through a corporate communication network.



— HEIDMANN - SLIDE 13 —

A couple of things. First of all, some video components I'd just like to mention. I'm sure some of you out there have had experience with some of these. In the basic computer graphics video system you've got the computer, a frame buffer and a monitor running in high resolution. Those are the areas in gray up on top, sort of covered by that little scum that I wiped on before I came in here this morning.

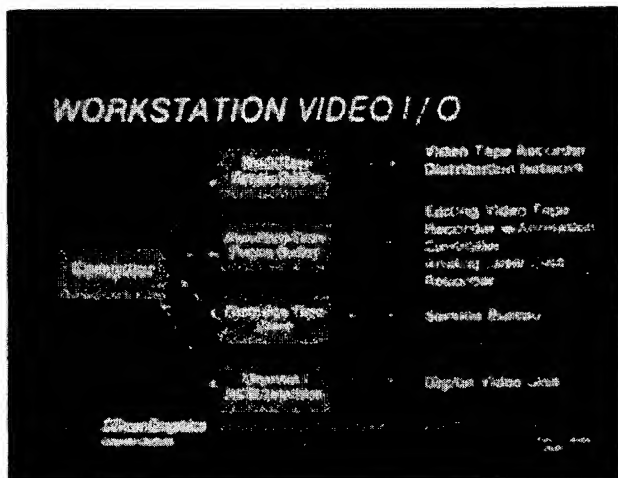
The first tool prevalent in a lot of computer systems is having a background low res frame buffer. That is, a buffer that sits on the system bus, but that works in NTSC or PAL to grab images or put images out there. It lets you deal with video images offline so you're not messing up what you're doing on your main screen. It lets you grab and display usually single frame at a time images, but if you've got the right recording equipment, that's fine.

The second thing is some sort of pixel blaster hardware that is still a separate frame buffer that talks NTSC or PAL, but deals with a high res frame buffer. If you're doing real time computer graphics you can send it out or bring it in and incorporate it with what you're doing. Much more complicated engineering arrangement, much more expensive, but a wider range of applications.

A Scan Converter is another way to get out to the video world. A Scan Converter is kind of a one-way path from your high res frame buffer out to video. It basically filters down a 1,000-line image to the 500-line image that you want for video in real time. There are a number of companies that make it. I believe HP has even integrated a board into their workstation.

Finally, something like a window keyer which allows you to take video from the outside world, combine it with your high res graphics and display it on your monitor. So that's a way of incorporating video into the work that you're doing on your monitor. Again, machines like this are available from a couple

of different companies. RGB Technologies is one that comes to mind.



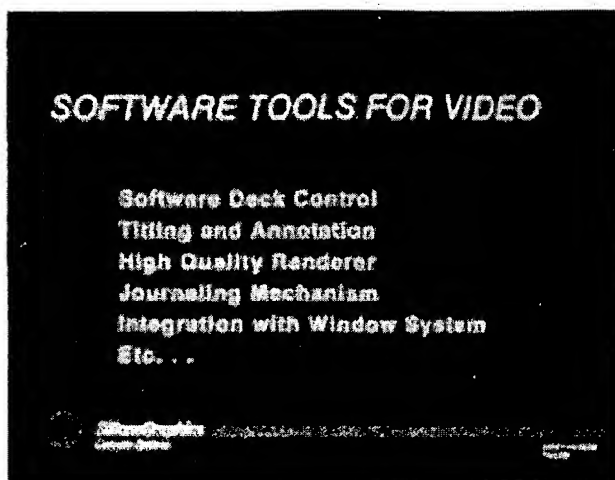
— HEIDMANN - SLIDE 14 —

In the area of output you've also got a few options here. The first thing, if you've got a real time frame buffer where you can produce images -- usable scientific or engineering images in real time -- you just need to go out to a videotape recorder or distribution network. You may need an encoder to go from separate RGB signals into a single composite video signal. I'd like to see that be a standard part of a computer workstation. You should just have one tap where you plug the cable in and it goes out and it's video that people can look at.

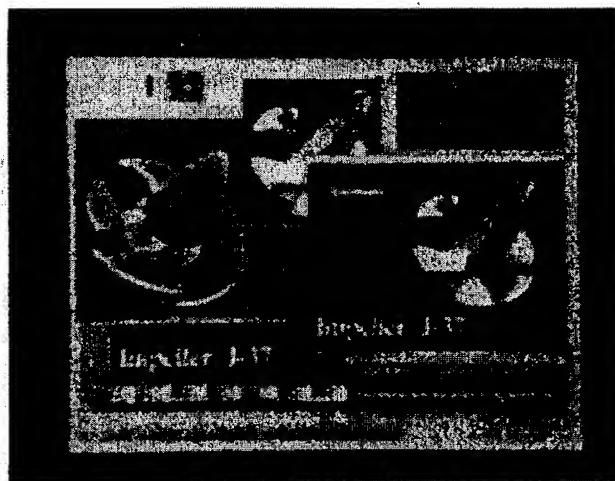
If you've got a non-realtime frame buffer it gets a little more complicated. You need a special videotape recorder that has editing capabilities plus something to control it. This is where wouldn't it be nice if that was just one piece of equipment. I think some people would buy that.

Another option is the analog laser disc. Panasonic, for instance, makes a laser disc recorder. It's in the area of \$10,000 to \$12,000 I think. That takes composite video in, and an RS232 signal, and you can lay frames down really quickly onto it and then play it back in real time. This is a really nice solution. As I said, for making nonrealtime real time, it's very fast, very convenient.

Another option that I think should become more important as time goes by is the ability to take a cartridge tape, send it someplace and get nice video back. (Applause) Somebody is looking for that out there I guess. I don't think there's anything established right now, but that's going to happen in a couple of years -- within the next couple of years. Finally, there are some video devices that will sit on the computer network.



— HEIDMANN - SLIDE 15 —



— HEIDMANN - SLIDE 16 —

Just one final note before I go. Good video on workstations isn't just video equipment. You also need software to go with it. Just about every video application needs some sort of titling capabilities. Deck control from the workstation. You'd like to have it work within the windowing system, so if you've got an application that's putting frames up there, but maybe not in real time, the windowing system should be smart enough to realize when a new frame comes up and tell the animation recorder to fire that.

Again, having a high quality renderer so that even if you're just an engineer, working where really all you care about is the wire frame, at some point in the process you might like to make a nice high quality picture, and I think we're seeing that from a lot of different vendors. So that's the direction we're going in.

That's all I have to talk about. We'll have a chance to do some questions later on, and if there's anybody out there making these products, let's go.

Next I'd like to introduce Floyd Wray. Floyd is a videographer and employee of the BYTE-by-BYTE

Corporation, which manufactures the Sculpt 4D package for the Amiga and now the Macintosh. Floyd.

**Floyd Wray**  
**BYTE-by-BYTE Corporation**

Good morning. Since Tim, Michael and Greg and I first got together to plan our attack, two things have become clear. First, that this panel was going to take place entirely too early on a Friday morning, and second, in spite of that fact, I have looked forward to what they had to say on the subject. I knew it would be enjoyable and informative -- at least right up to this point, right here, where it became my turn to talk.

This has to be the worst part of my day. In some ways my angle on the subject almost comes in as a summary for what has been said, and it's not an obvious summary at all because it deals with how to go about buying desktop gear.

Most of us will probably make a desktop purchase in the not-too-distant future if we haven't already. What do you buy first and why? How do you model the future of desktop video as it relates to purchasing?

Now if those questions aren't tough enough, how do you further explain obsolescence upon purchase to a banker? For as you know, the minute you invest in desktop technology, you just bought an antique. As has been said, one of the first things that comes to mind when you mention desktop anything is desktop publishing. Unfortunately, a pretty good argument could be made that desktop publishing was an accident -- like Columbus, who accidentally discovered the Americas. Computer manufacturers charted a course toward ease of use and power and ended up discovering desktop publishing enroute. And here is the distinction.

With desktop publishing we discovered the capability first. Then we named it. With desktop video, in the grand tradition of American marketeering, we discovered the name first. As a result, desktop video tends to be in the words of St. Paul -- all things to all men.

A first order of business involves defining the subject. Notice how that theme keeps coming up. What is desktop video? Now your definition might be based on cost, hardware configuration, software. The point is, without some deep introspection at the definition end of this subject, it is impossible to make an intelligent purchase. I have a weird metaphor in this regard. Imagine two kingdoms residing atop two made-in-one-hour toxic green 3-D mountains. In the resulting valley these two kingdoms share a vast technological border. On one side is the Kingdom of Video, populated by a sturdy race. Perhaps we should point out that the inhabitants here are largely analog in nature.

On the other is the Kingdom of the Computer, populated by an equally competent race, full of answers, restless, innovative, and by contrast, incredibly digital. Now depending upon your degree of cynicism, desktop video is either the land bridge between these two kingdoms or the landfill. And that gives us a pretty good perspective. The technology being tossed into this gap by companies on both sides of the boundary -- with this simple metaphor of two kingdoms we have the basis for a purchasing strategy.

The next question is equally basic. What is the chief enterprise in desktop video? Desktop video deals ultimately in what might be called audiovisual sentencing. On the videographer side of the gene pool, we use cameras to trap bulky visual sentences. Once we've stored them on tape, we

send these sentences to a fat farm, also known as the edit suite, where we trim them and edit them together.

On the other side of the gene pool the computerists tend to author visual sentences. Where the camera traps images, computers are used to construct them. The video desktop is thus involved in two things -- the enterprise of audiovisual trapping, and the enterprise of audiovisual authoring. All desktop gear can be lumped into one category or the other.

After the problem of basic definition, one of the toughest issues regards where to go for purchasing counsel. Now videographers can give you pointers on formats and sync configurations. Computerists of course can explain the CPU and frame buffers and criteria for good software. Chances are though most desktop video buyers are looking for technology that exists along the outside edge of either specialty, and if you probe either side in an effort to stimulate a little wisdom on the subject, you'll also most likely hear faint traces of good old fashioned technological racism.

For example, I'm going to do a little characterization. If you crept over the border and listened to videographers around the campfire at night you might hear something like this: Computer weenies honestly think they're going to change the world with two super VHS machines -- not even three-quarter -- and \$50,000 editing software. Someone ought to take a missionary journey over there and help them drag their production values out of the mud. Computerists of course can also demonstrate their own version of bigotry. Around their campfires you hear: All a video weenie thinks about are character generators and flying logos. Ha, ha, ha. When the day of digital fully dawns, we will blow them to the other side of the universe. Hold onto your Old World shabby NTSC pants, folks; help is on the way. (Applause)

I've heard this pretty much. This duality is critical to your purchase because both sides have information you need. They also may have some misinformation. You are the one who makes the call based upon exactly what you expect from your desktop. So solid definition of desktop video according to your own expectation is a first order of business. Second, it's important to generate an equipment list based on as much information as you can dig up. And third, we must deal with those dark and wonderful prophecies of the future -- those technologies that bear upon our decision-making because they're just around the corner. And that takes us back to that earlier issue of obsolescence.

All too often some of us get so hamstrung in information gathering and what's just around the corner, that we end up at the polar extreme -- not doing anything in the present.

I guess this points up one of the few known quantities on the subject. Desktop videographers are spending their own money and maybe this is a powerful clue for our ongoing definition. Maybe we should try to ignore describing desktop video in terms of price and power. Maybe desktop video really should be -- as has been said -- mostly defined as an enterprise of the individual. We may also discover that desktop video is not as much technology as it is human passion -- driving us forward to a new form of literacy.

Now in my own notion of the future, I see emerging 3-D technologies as vital to any purchase. With this in mind, I want to show three animations. Unlike demos produced in a huge production setting, each animation here was animated by small team -- one or two people at the most -- on PCs.

— VIDEOTAPE PLAYING —

But at SIGGRAPH we expect high resolutions. High resolutions. That's how it should be. But there's another resolution we should be concerned with -- individual resolution. I don't believe that desktop video is the frail brother of desktop publishing. Every desktop computer consumed with word processing today is a potential site for tomorrow's audiovisual authoring. With so much happening, with so much change, how do you plot a course? How do you resolve your final purchase dilemma in the face of obsolescence?

Well, if desktop video truly is an enterprise of the individual, then obsolescence may be an appropriate description for those of us who never get involved for the fear of investing in the wrong thing. As it turns out, desktop videographers are not investing in technology. Desktop videographers are investing in themselves. They buy, they close their eyes to new stuff that comes out the next day, and they content themselves to produce with the hardware that they have.

Now, they may eventually suffer an attack of pride as they work with their antiques, their old cows, but they learn. They learn to milk those old cows until they're dry. Crackfish was assembled on a Sony 5850, which I bought in 1985. It wasn't Betacam SP or D1 or D2 or even three-quarter SP. I love my old cow.

Let us rewrite our understanding of what's actually happening here. Desktop video could probably be achieved with a \$200 Gold Tongue VHS machine from Taiwan and a UHF transmission. The point is after you have achieved a bit of clarity and spent an appropriate time researching equipment -- jump in. When it comes to your future desktop purchase, do not fear most the buying. Fear most the not doing. Thank you.

Moderator  
Tim Heidmann  
Silicon Graphics

We're going to try and go to these microphones. Am I audible? We've got a few minutes to do some questions and answers. We've got microphones set up around the hall. If you'd like to bring up a point, contradict anybody, please just step to a microphone and make yourself known.

Q. You were talking about the different types of equipment that might be needed to take a computer signal and put it on videotape. Do all of the PCs that we were looking at -- we were looking at Silicon Graphics equipment, we were looking at the Amiga, I think we were looking at like IBMs or something like that. Do they all have that same type of problem? Are there any computers that it's less of a hurdle by the nature of the design of the computer, or is it just the same standard problem with each of those computers?

HEIDMANN: Everybody stinks. Do you have anything to say, Michael on that?

MacKAY: Yes, I've actually had quite a bit of experience being the one that has to come in the room and make this thing go on the tape. So basically the main problem is that video has a very defined specification -- defined by the EIA -- Electronic Industries Association. We've entered into a world where everybody wants to do single frame editing and this puts even tighter constraints on there. Companies like Silicon Graphics have done a good job of allowing you to go into NTSC mode. The only problem is you can't see your other monitor. So it makes it hard to get back to UNIX. Computers like the Amiga, they have RS170A style -- actually RS170 -- style video outputs, and they actually work quite well, and if

you find in the PCs and Macintosh style platforms that actually have display adapters that can be added in after the fact, if you do some homework and actually find some stuff, there's a lot of products that are very high quality, that by choosing the correct display adapter can produce very high quality video.

Then it's just more of decisions about dealing with synchronization, like Greg pointed out, and being able to genlock that. So depending on your application, I'd say go with the computing platform that has the most flexibility in choosing display adapters. You cannot record VGA cards and you cannot record 1125 60 displays, like the 1280 by 1024 -- 60 hertz stuff that's the typical console on a high end workstation. But there are a lot of cards and I'm not going to try and name people right now. But if you do some research, and I'd be glad to give you a list of some of these things after. HEIDMANN: So, no easy answers. One good idea is always if you're thinking about some hardware, find somebody who's using it, somebody hopefully who's doing the same sort of thing you hope to do, and talk with them. Because you're going to find troubles doing something with everything. So just pick the piece of equipment that is at least you know is good for doing what you want to do. Yes.

Q. I had a question for Michael. You mentioned the high band 8 millimeter. I'm not familiar with that term. I have a Sony CCD V9 8 millimeter camera; it's a great camera. I love it. Is the high band tapes -- can they be used in that camera? And how is the high band different from a regular 8 millimeter format?

MacKAY: What goes on is that there are incremental improvements in technologies that have come about -- mostly actually from tape formulations is what we're seeing right now -- where now they're actually using some of the techniques used to manufacture semi conductors in the manufacturing of videotapes. So what we're getting is tapes with higher coercivities, and without going into all the terms about the manufacturing of tape, it allows you to do higher bandwidth recording. So what we're doing is we're pushing up the carrier frequencies in these decks. U-matics have taken a step forward, and there is now what's called U-matic SP, which is called Sony's Superior Performance SP U-matic, and basically that same technology has been applied to the Betacam line and is being applied to now the 8 millimeter line. I own a V9 myself. They're all upward compatible. In other words, if you buy an SP three-quarter inch or a high band 8 millimeter or even the still image stuff, the older medium will stay play in it. You won't take advantage of the format. But basically high band 8 millimeter is over 400 lines. Sony is also very conservative on their specifications, and it has a great improvement in chrominance signal to noise ratio, with the difference being in S-VHS they have not made any of these improvements in the chrominance area. They have all been in the luminance area, and we like color pictures. So you need both.

MacNICOL: I'd like just to add that while a lot of these formats are really good, the key element in creating a system is integration. For instance, while high-band 8 is very high quality, you have to look at what interface controllers will work with it, and also which ones will work with frame accuracy.

For instance, VHS, which is good -- S-VHS, which is better -- in spite of its qualities, sometimes has difficulty getting still frame accuracy. So if you create an animation with that, occasionally you'll have a frame that's missing or a frame that's extra. This is the reason why if you are creating a low

cost animation, you really have to look at the issue of compatibility.

**HEIDMANN:** We have a question over on my left.

**Q:** I wanted to thank the speakers for a great set of talks. Then I wanted to point out an area that they hadn't discussed. One of the phenomenon in computer use today is the use of computers for interpersonal communication online, and no one's mentioned the possibilities for video enhancing that interpersonal communication.

**HEIDMANN:** Certainly a big application area which we haven't talked about much, which I guess falls into the distribution and viewing area. Again, the hardware manufacturers, what hooks are we making to make that easy to get video on a computer screen -- if that's the way you want to do it. And how do you make effective use of that in software, so that you're not just setting up a camera and running a cord over to a monitor and then setting up a camera and running a cord over to the monitor. So the writing, coding and compression make use of that. There is a lot of stuff we didn't cover today -- in case anyone was wondering. Yes.

**Q:** As we have learned with film, tape doesn't last very long. Is there a way to preserve a final product so that we can see it 10, 20, 30 years from now?

**HEIDMANN:** Anyone volunteer to answer that?

**MacKAY:** Basically what you do is you rerecord tapes periodically -- anybody that's dealing with the large archiving things. And if you have the need to archive something over that period of time, then you really need to preserve a digital archive of the original byte maps of the renderers in some format. There has been a lot of third party development using 8 millimeter as a terabyte storage medium, and being able to use mass quantities of stuff. Don't tell anybody, but we're trying to develop -- I have a proposal on the table right now to do a hybrid 8 millimeter deck that's also a streaming tape drive and a single frame editable animation controller -- all in the same thing.

**HEIDMANN:** There's a question in that corner.

**Q:** I'd just like to describe a situation that we're working in and see if you know people that are working on it or if you are. We built a digital video studio in CBC in Toronto and from what I've seen here, in a sense you are talking about simulating what already people do in Vidcoland -- basically Analogland that's becoming more digital. But from a production point of view, I'm wondering if people are working on how you prepare for complex productions because what we see on the floor with desktop video is basically simulating what an operator does when they're editing. But when you're dealing with complex ideas -- for example, in a documentary. You want to go into a postproduction suite where you have mega layers of materials -- all with different virtual points of view, all with different kinds of timing on them. There is no way that you can pre-view that now. You're working with paper scripts and awkward story boards. And I haven't seen anything in effect. Computer supports the intellectual production part of the production, rather than just the technical part of how you plug in machines, as digital and video get together. I think that's really where we talk about what we're actually saying with the images -- whether they're generated or whether they're captured. There's a language that's been developed in the images. And if we don't look at that issue as we move into the computer world, we won't have those tools either.

**HEIDMANN:** Greg, did you want to say something?

**MacNICOL:** That's a really good point. That really is excellent. Recently I wrote an article just focusing on that

issue. When I wrote the article it was -- typically there's about a two to three month lead time, and by the time I had the article out, everything was completely different. I focused on about two companies focusing on that issue -- editing, video editing. And at the moment there are about at least five companies -- and I know about five other companies -- that are developing very powerful and complex video editors.

The wonderful feature about these is that instead of requiring about a week of training on a very advanced video editing system, it's much easier to use, and it's based not on rows and rows of numbers and time code, but pictures -- the cut in and the cut out point. So this is a very serious issue and this is also how all low cost systems -- in fact, we're seeing now on low cost Amigas and of course on Macs -- we're seeing these systems being used to replace \$100,000 editors. Now these systems aren't complete yet. Some of these systems are not frame accurate; some of them are close to three or four frames, which for a professional studio is not good enough -- and especially for computer animation -- is not good enough. But I think in the next year we're going to see some very impressive developments.

**HEIDMANN:** Thank you. Unfortunately, I've been told we've run over time and we won't be able to take any more questions in this forum. We have to end the panel. But the panelists will be staying around. If you do have some more questions you'd like to follow up on, please come on up. We can continue this out in the hall perhaps.

Thank you very much for your attendance.

## **ELEFANT & KANGURU: HARDWARE CONTROL PANELS DEDICATED TO DISK SERVERS FOR SLOW MOTION, EDITING AND MULTIPLAY APPLICATIONS**

J. L. Grimaldi and M. Bougard

Numeric Video, France

### **ABSTRACT**

The advent of video disk servers has brought solutions to the important limitations of tape recorders which are:

- replay while recording by use of multi channel I/O and sharing of storage resources
- virtual editing

The graphics user interfaces (GUI) provided with the video disk servers, if they appear very attractive at first glance, with graphics, pop up menus, dialog boxes etc, become difficult to use in events like sports and news in which a short response time is one of the major characteristics. Thus, the dramatic improvement of productivity in 'no delay' operations such as sports coverage when using video servers, emphasized the absolute necessity to have modern control panels able to exploit all virtual edit features, providing fast response time and easy ergonomics. Similarly, virtual editing and sharing of recorded material among multiple outputs allowed the development of Clip bank systems (a system in which virtual edits are accessible in any order, at any time on any output).

Developed with and for the operators, Elephant and Kanguru provide all slow motion and edit functionalities such as Sequence and Clip creation, storage of Edit Decision lists, easy update of these lists as well as change of Edit entry points, multi record capability, Time delay, ...

In the paper, the authors will describe all major features of modern dedicated control panels such as Elephant. Emphasis will be put on user requirements and important productivity tools provided by Elephant. A comprehensive description of typical operating sessions (Sequence creation, Editing, ...) will be explained.

### **OVERVIEW**

The advent of video disk servers has brought solutions to the important limitations of tape recorders which are:

- replay while recording by use of multi channel I/O and sharing of storage resources,
- virtual editing,
- multiple replays on different outputs at different time of the same video material.

These new facilities have helped fast and cost effective development of slow motion for sports and on the fly, news edit. However to remain cost effective the industry asks for a significative

improvement of productivity in these 'no delay' operations such as sports report and news edit. It leads to the absolute necessity of having modern control panels providing fast response time and easy ergonomics, able to exploit all virtual edit features.

Most video servers propose, as a basic feature, a graphics user interface (GUI) which is in some way a low cost and quite attractive solution to start with. After training the GUI, despite graphics, pop up menus, dialog boxes etc, is difficult to use on occasions in which a short response time is one of the major needs.

It is to meet the requirements of fast and efficient control panels that Elephant and Kanguru have been developed.



In the following, we will highlight the main points of our development starting with User Requirements, continuing with Numeric Video philosophy of Control panels and, as a conclusion, detailing some specific features.

## USER REQUIREMENTS

The goal of the study was to define as precisely as possible user requirements trying to distinguish absolute requests from useful features.

We were conscious that it is impossible to meet all requests with a single machine and we also tried to identify different segments of activities. The task was also complicated by the fact that different types of video servers exist. Since we had the objective to make a somewhat high-end product with a maximum of features, we chose a class of video server with multiple I/O and providing a level of performances matching our objectives.

During the set up of the questionnaire, we created a model of one video server with 4 Channels (either Input or Output) and one control panel. This model was somewhat arbitrary, based on existing pieces of equipment available on the market.

We started with one application : slow motion for sport. With this in mind, we asked :

- how many records had to take place simultaneously ;
- how much recording time was necessary and for which quality if in a compressed system ;
- how many video and audio I/Os were necessary ;
- how compact the control panel should be ;
- describe or imagine an operating mode for slow motion and by consequence what are the needed tools, buttons etc... to obtain the best response time ;
- should multiple control panels run concurrently on the same server ;
- how autonomous should be the control panel (had it to replace configuration tools, disk management, ...) ;
- what other application would you like to run on the same control panel.

The questionnaire was sent to operators, OB VAN technical managers and directors. Conclusions of the surveys were :

- It is highly desirable that by using adequate control panel, a single operator can operate the 4 channels, setting up 3 in Record, one in Replay. This saves space and operating costs.
- The system has to be flexible enough to allow any kind of combination between 1 and 3 simultaneous recorders, with changes done on the fly.
- The control panel must also be used to configure the video file server and avoid as much as possible the use of computer monitor, keyboard and mouse.
- The same operator during the same work session on the same control panel should be able to realize Instant Replay (slow motion) as well as preparing Highlight (a virtual Edit made of an assembly of Sequences). The possibility offered by video servers to keep recording while Editing is a big feature.
- Control panel must be powerful and be easy to operate (at least for the basic functions Replay, Speed Control, Mark In, ...)
- Control panel must be compact, have the right number of keys to minimize the number of keystroke (50% of action with one or 2 keystrokes) and display information (TimeCode, Replay Speed, Edit List, ...) in a very ergonomics manner. Jog/Shuttle and T-Bar are mandatory devices.
- To have a Clip Store application and possibly a Time Delay with Editing capabilities was also a request.
- To dispose of a low level application to independantly manage the 4 channels.
- Training time must be minimized.

## NUMERIC VIDEO APPROACH

We began by sorting out the important points of the User requirements in terms of impact on Control Panel design . Here is the list of the most significant items :

- Size : As operator works facing the equipment : should be smaller than a body width, should minimize fingers' movements, must provide a sufficient number of buttons and accessories



- Form : be a nice object (operators will work with for long hours, e.g. tennis tournament)
- General Ergonomics : Buttons must be lit, readings be clearly visible in a dark ambience, Jog/Shuttle and T-Bar easy to grasp and handle.. Display must be easy to read (good contrast, large body police)
- Operating system 'transparent' : operators do not want to manage files, exception errors, ...
- Have sufficient memory to handle a large number of Edits

Considering that we received many requirements from the users, it was obvious that we could not satisfy all with a single application. In a second phase, we tried to split into different applications while keeping the same Hardware. The idea was that doing so we would be able to cover a wide range of applications by just changing Software. Moreover, we decided that we will also offer a solution covering all applications .

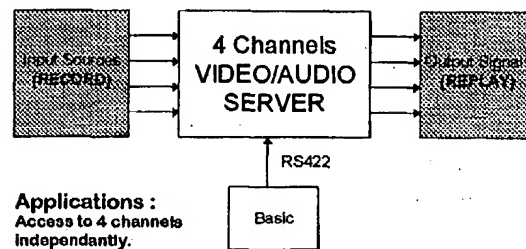
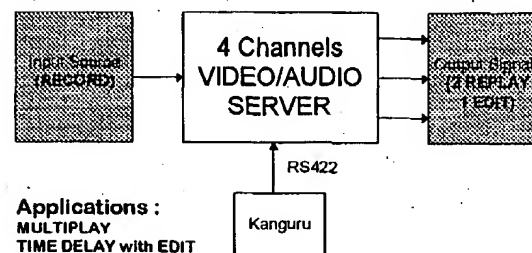
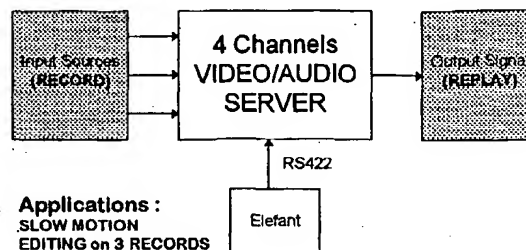
#### IDENTIFIED APPLICATIONS

As foreseen, the first application is **SLOW MOTION and EDIT**, which applies to Sports or News Edit. It implies multiple Recorders and one channel devoted to Replay and Editing. This application minimizes the access to all Recorded materials.

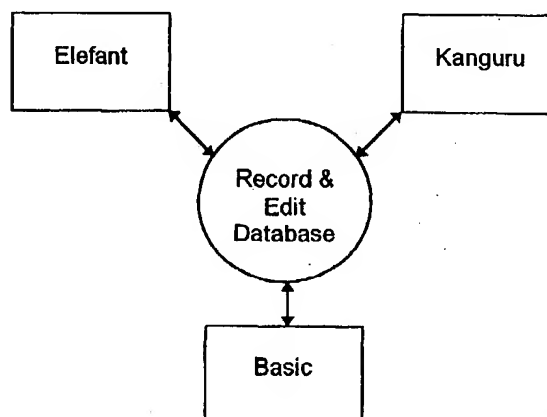
Interestingly, the dual application (in the sense that it reverses the use of Player and Recorder) is **MULTIPLAY**, in which 3 channels are Players (2 for Program Outputs of Clips or Records, One for Edit), one is a Recorder (Time Delay). This application provides very efficient access to all Edits.

The third 'application', **BASIC**, allows to use the different channels individually as Recorder or Player.

The 2 main applications have been named respectively **Elefant** (emphasizing its wide memory of Edits) and **Kanguru** (emphasizing its ability to jump from clip to clip). Basic is indeed an embedded part of both and thus did not receive a specific name.



All applications share a common database, which means that Edits made with one application can be used by others. For example Highlights (summary) of a sport event made with Elefant can be re-used later in Multiplay (Kanguru).



## PRACTICAL IMPLEMENTATION OF EDIT FUNCTIONALITIES

It is a strong and legitimate request of users that Edit functionalities have to be as easy and fast as possible. This request is very well addressed into Elephant and Kanguru.

By definition, a Clip is a virtual EDIT made of an EDIT LIST : ordered list of Sequences (MarkIn, MarkOut on a Record). The EDIT list is displayed on the graphics screen of Elephant or Kanguru under the following form :

<b>START</b>		
00:00:00.00	CUE	801
00:00:01.00	CUE	802
00:00:03.05	SEP	0005m
00:00:06.20	SEQ	100
<b>END</b>		

Either in Elephant or Kanguru products, all EDIT keys functions are grouped as a block of 6.

INS	MOVE	EDIT
SUPP	ESC	STORE

- **EDIT : FORCES ENTRY INTO EDIT MODE**

Since Elephant or Kanguru have different operating modes, EDIT key is used to enter into EDIT mode, i.e. to create CLIPS.

- **INS : ADDS ELEMENTS TO EDIT LIST**

SEQUENCES and CLIPS can be inserted to the EDIT LIST using 'point and pick' cursor .

- **SUPP : REMOVES ELEMENTS FROM EDIT LIST**

- **MOVE : CHANGES ORDER OF ELEMENTS**

- **STORE : STORES EDIT LIST IN MEMORY**

Additional functions, such as modifying In and Out points of each SEQUENCE of a CLIP, within the EDIT LIST are provided. The direct availability of all these functions have been plebiscited by users since they minimize the time for Editing and provide a very powerful interface. Highlights are realized very quickly while recording live sources. For example, a Clip of 4 Elements is created with the following Keystrokes :

1. EDIT (recalls existing Clip or define a New one)
  2. INS (4 times) : Picks into the ELEMENT LIST and copies into EDIT List
  3. ESC (to exit from INS sub mode)
  4. STORE (stores current Edit list in memory)
- 7 Keystrokes are sufficient to create a virtual Edit of 4 elements. This can be done in less than 30 seconds !!!

## CONCLUSIONS

Close relationship with the operators have led to the development of a modern control panel providing easy access to Editing and Clips. Elephant and Kanguru have been well received by user here and in the USA and have launched a new way of operating Virtual Edit equipment.